

EXPRESS MAIL CERTIFICATE

44 Rec'd PCT/PTO 2 5 SEP 2000

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. § 1.10 on the date indicated above and is addressed to the Assistant Commissioner for Patents, Washington, D.C. 20231 on this date: 25 September 2000

Typed or Printed Name	Dave Glisson	Express Mail No.	EL539 356 278US
Signature	Dave Glisson	Date	25 September 2000

Form PTO-1399
(REV 10-94)

U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE

ATTORNEY'S DOCKET NUMBER
STAN-084

TRANSMITTAL LETTER TO THE UNITED STATES
DESIGNATED/ELECTED OFFICE (DO/EO/US)
CONCERNING A FILING UNDER 35 U.S.C. 371

U.S. APPLICATION NO.

097647067

INTERNATIONAL APPLICATION NO.
PCT/US99/06573INTERNATIONAL FILING DATE
25 March 1999PRIORITY DATE CLAIMED
26 March 1998

TITLE OF INVENTION
NOVEL MAMMALIAN G-PROTEIN COUPLED RECEPTORS HAVING EXTRACELLULAR LEUCINE RICH REPEAT REGIONS

APPLICANT(S) FOR DO/EO/US
HSUEH, Aaron J.W.; HSU, Sheau Yu; LIANG, Shan-Guang; VAN DER SPEK, Petrus Johannes

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☒ This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. ☒ This express request to begin examination procedures (35 U.S.C. 371(f)) at any time rather than delay examination until the expiration of the applicable time limit set in 35 U.S.C. 371(b) and the PCT Articles 22 and 39(1).
4. ☒ A proper Demand for International Preliminary Examination was made by the 19th month from the earliest claimed priority date.
5. ☒ A copy of the International Application as filed (35 U.S.C. 371(c)(2)):
 - a. ☐ is transmitted herewith (required only if not transmitted by the International Bureau).
 - b. ☐ has been transmitted by the International Bureau.
 - c. ☒ is not required, as the application was filed in the United States Receiving Office (RO/US).
6. ☐ A translation of the International Application into English (35 U.S.C. 371(c)(2)).
7. ☒ Amendments to the claims of the International Application under PCT Article 19 (37 U.S.C. 371(c)(3)):
 - a. ☐ are transmitted herewith (required only if not transmitted by the International Bureau).
 - b. ☐ have been transmitted by the International Bureau.
 - c. ☐ have not been made; however, the time limit for making such amendments has NOT expired.
 - d. ☒ have not been made and will not be made.
8. ☐ A translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).
9. ☒ An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).
10. ☐ A translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

Items 11 to 16 below concern other document(s) or information included:

11. ☐ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
12. ☒ An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.
13. ☐ A **FIRST** preliminary amendment.
14. ☐ A **SECOND** or **SUBSEQUENT** preliminary amendment.
15. ☐ A substitute specification.
16. ☒ A change of power of attorney and/or address letter.
17. ☒ Other items or information: postcard

097647067

534 Rec'd PCT/PTO 25 SEP 2000

17. ☒ The following fees are submitted:
Basic National Fee (37 CFR 1.492(a)(1)-(5)):

Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and international search report not prepared by the EPO or JPO \$970.00

International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search report prepared by the EPO or JPO \$840.00

International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO \$690.00

International preliminary examination fee (37 CFR 1.482) paid to USPTO but all claims did not satisfy provisions of PCT Article 33(1)-(4) \$670.00

International preliminary examination fee (37 CFR 1.482) paid to USPTO and all claims satisfied provisions of PCT Article 33(1)-(4) \$ 96.00

ENTER APPROPRIATE BASIC FEE AMOUNT = \$ 670

Surcharge of **\$130.00** for furnishing the oath or declaration later than _____ 20 ____ 30 months from the earliest claimed priority date (37 CFR 1.492(e)).

\$

CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE		
Total Claims	18 - 20 =	-0-	X \$18.00	\$	-0-
Independent Claims	8 - 3 =	5	X \$78.00	\$	390
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			+ \$260.00	\$	-0-

TOTAL OF ABOVE CALCULATIONS = \$ 1,060

Reduction of $\frac{1}{2}$ for filing by small entity, if applicable. A Small Entity Statement must also be filed.
 (Note 37 CFR 1.9, 1.27, 1.28)

\$ -0-

SUBTOTAL = \$ 1,060

Processing Fee of **\$130.00** for furnishing the English translation later than _____ 20 ____ 30 months from the earliest claimed priority date (37 CFR 1.492(f)).

\$ -0-

TOTAL NATIONAL FEE = \$ 1,060

Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). \$40.00 per property.

\$ 80

TOTAL FEES ENCLOSED = \$ 1,140

Amount to be
refunded \$

charged \$

a. ___ A check in the amount of \$___ to cover the above fees is included.

b. ☒ Please charge my Deposit Account No. 50-0815 in the amount of \$1,140. to cover the above fees.

c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 50-0815.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO

BOZICEVIC, FIELD & FRANCIS LLP
 200 Middlefield Road, Suite 200
 Menlo Park, CA 94025
 (650) 327-3400 Telephone
 (650) 327-3231 Facsimile

SIGNATURE

NAME Bret E. Field

37.620
 REGISTRATION NUMBER

NOVEL MAMMALIAN G-PROTEIN COUPLED RECEPTORS HAVING
EXTRACELLULAR LEUCINE RICH REPEAT REGIONS

INTRODUCTION

5 Field of the Invention

The field of this invention is the G-protein coupled receptor family of proteins.

Background

Gonadotropins (Luteinizing hormone, LH; follicle stimulating hormone, FSH; chorionic gonadotropin, CG) and thyrotropin (TSH)) are essential for the growth and
10 differentiation of gonads and thyroid gland, respectively. These glycoprotein hormones bind specific target cell receptors on the plasma membrane to activate the cAMP-protein kinase A pathway.

The receptors for LH, FSH and TSH belong to the large G-protein-coupled, seven-trans-membrane protein family but are unique in having a large N-terminal extra-cellular
15 (ecto-) domain containing leucine-rich repeats important for interaction with large glycoprotein ligands. Studies suggest that in these receptors, the extra-cellular leucine rich repeat region serves as a "baseball glove" which efficiently catches its corresponding large hormone ligand and optimally orients it for interaction with the seven trans-membrane-helical domain of the receptor.

20 Because hormones and receptors play a prominent role in a variety of physiological processes, there is continued interest in the identification of novel receptors and their ligands, as well as the genes encoding the same.

Relevant Literature

References of interest include: el Tayar, N, "Advances in the Molecular
25 Understanding of Gonadotropins-Receptors Interactions," Mol. Cell. Endocrinol. (December 20, 1996) 125: 65-70; Bhowmick et al., "Determination of Residues Important in Hormone Binding to the Extracellular Domain of the Luteinizing Hormone/Chorionic Gonadotropin Receptor by Site-Directed Mutagenesis and Modeling," Mol. Endocrinol. (September 1996) 10: 1147-1159; Thomas et al., "Mutational Analyses of the
30 Extracellular Domain of the Full-Length Lutropin/Choriogonadotropin Receptor Suggest Leucine-Rich Repeats 1-6 are Involved in Hormone Binding," Mol. Endocrinol. (June 1996) 10:760-768; Segaloff & Ascoli, "The Gonadotropin Receptors: Insights from the

Cloning of their cDNAs," Oxf. Rev. Reprod. Biol. (1992) 14: 141-168; Braun et al., "Amino-Terminal Leucine-Rich Repeats in Gonadotropin Receptors Determine Hormone Selectivity," EMBO J (July 1991) 10: 1885-1890; and Segaloff et al., "Structure of the Lutropin/Choriogonadotropin Receptor," Recent Prog. Horm. Res. (1990) 46: 261-301.

5

SUMMARY OF THE INVENTION

Three novel mammalian G-protein coupled receptors having extra-cellular leucine rich repeat domains, i.e. LGR4, LGR5 and LGR7, and polypeptide compositions related thereto, as well as nucleotide compositions encoding the same, are provided. The subject proteins, polypeptide and nucleic acid compositions find use in a variety of different applications, including the identification of homologous or related genes; the production of compositions that modulate the expression or function of the subject proteins; in the identification of endogenous ligands for the subject orphan receptors; in the generation of functional binding proteins for the neutralization of the actions of endogenous ligands; in gene therapy; in mapping functional regions of the protein; and in studying associated physiological pathways. In addition, modulation of the gene activity *in vivo* is used for prophylactic and therapeutic purposes, and the like.

10

15

BRIEF DESCRIPTION OF THE FIGURES

20

Fig. 1 provides the nucleotide and amino acid sequence for human LGR4.

Fig. 2 provides the nucleotide and amino acid sequence for human LGR5.

Fig. 3 provides the nucleotide and amino acid sequence for human LGR7, long form.

25

Fig. 4 provides the nucleotide and amino acid sequence for human LGR7, short form.

Fig. 5 provides an alignment comparison of the long and short forms of LGR7.

Figs. 6 provides a comparison of deduced amino acid sequence of LGR4 and 5 cDNAs and those encoding FSH and LH receptors.

30

DESCRIPTION OF THE SPECIFIC EMBODIMENTS

Novel mammalian G-protein coupled receptors having extra-cellular leucine rich repeat regions (i.e. LGR4, LGR5 and LGR7) and polypeptide compositions related thereto, as well as nucleic acid compositions encoding the same, are provided. The

5 subject polypeptide and/or nucleic acid compositions find use in a variety of different applications, including the identification of homologous or related genes; for the identification of endogenous ligands for these novel receptors; the production of compositions that modulate the expression or function of the receptors; for gene therapy; for mapping functional regions of the receptors; in studying associated physiological

10 pathways; for *in vivo* prophylactic and therapeutic purposes; as immunogens for producing antibodies; in screening for biologically active agents; and the like.

Before the subject invention is further described, it is to be understood that the invention is not limited to the particular embodiments of the invention described below, as

15 variations of the particular embodiments may be made and still fall within the scope of the appended claims. It is also to be understood that the terminology employed is for the purpose of describing particular embodiments, and is not intended to be limiting. Instead, the scope of the present invention will be established by the appended claims.

20 In this specification and the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this invention belongs.

25 CHARACTERIZATION OF LGR4, LGR5 AND LGR7

LGR4, LGR5 and LGR7 are novel mammalian receptors of the G-protein coupled, seven trans-membrane family of proteins, specifically the subfamily of G-protein coupled seven trans-membrane proteins which are characterized by the presence of extra-cellular leucine rich repeat regions. As such, these proteins have trans-membrane segments and

30 extra-cellular regions similar to those found in the known LH, FSH, and TSH receptors. In other words, these proteins have both a G-protein coupled seven trans-membrane region

and a leucine rich repeat extra-cellular domain. The N-terminal extra-cellular domains of these proteins also show high homology with Drosophila Slit and Toll proteins having leucine rich repeats. These proteins are expressed in diverse tissues.

The human LGR4 gene has a nucleotide sequence as shown in SEQ ID NO:01.

- 5 The human LGR4 gene product has an amino acid sequence as shown in SEQ ID NO:02. LGR4 is expressed in a plurality of different tissue types, including ovary, testis, adrenal, placenta, liver, kidney and intestine.

The human LGR5 gene has a nucleotide sequence as shown in SEQ ID NO:03.

- The LGR5 gene product has an amino acid sequence as shown in SEQ ID NO:04. LGR5
10 has been found to be mainly expressed in muscle, placenta and spinal cord tissue.

- The human LGR7 gene encodes multiple splicing variants, each of which contains a multitude of cysteine-rich low density lipoprotein (LDL) binding motifs at the N-terminus in addition to the leucine rich repeat region. The longer forms of LGR-7 have a higher similarity than shorter forms of LGR-7 to snail LGR in the trans-membrane domain
15 and the N-terminal LDL binding domain. The overall structure of both the long and short forms of LGR-7 is similar to that of the LH receptor. The human LGR7 short form gene has a nucleotide sequence as shown in SEQ ID NO:05. The LGR7 short form gene product has an amino acid sequence as shown in SEQ ID NO:06. The human LGR7 long form gene has a nucleotide sequence as shown in SEQ ID NO:07. The LGR7 long form
20 gene product has an amino acid sequence as shown in SEQ ID NO:08. LGR7 is expressed in multiple tissues, including testis, ovary, prostate, intestine and colon.

IDENTIFICATION OF *LGR4*, *LGR5* AND *LGR7* SEQUENCES

- Homologs of *LGR4*, *LGR5* and *LGR7* are identified by any of a number of
25 methods. A fragment of the provided cDNA may be used as a hybridization probe against a cDNA library from the target organism of interest, where low stringency conditions are used. The probe may be a large fragment, or one or more short degenerate primers.

- Nucleic acids having sequence similarity are detected by hybridization under low stringency conditions, for example, at 50°C and 6×SSC (0.9 M sodium chloride/0.09 M
30 sodium citrate) and remain bound when subjected to washing at 55°C in 1×SSC (0.15 M sodium chloride/0.015 M sodium citrate). Sequence identity may be determined by

hybridization under stringent conditions, for example, at 50°C or higher and 0.1×SSC (15 mM sodium chloride/0.15 mM sodium citrate). Nucleic acids having a region of substantial identity to the provided *LGR4*, *LGR5* and/or *LGR7* sequences, e.g. allelic variants, genetically altered versions of the gene, etc., bind to the provided sequences

- 5 under stringent hybridization conditions. By using probes, particularly labeled probes of DNA sequences, one can isolate homologous or related genes. The source of homologous genes may be any species, e.g., primate species, particularly human; rodents, such as rats and mice; canines; felines; bovines; ovines; equines; yeast; nematodes; etc.

- Between mammalian species, e.g., human and mouse, homologs have substantial
10 sequence similarity, e.g. at least 75% sequence identity, usually at least 90%, more usually at least 95% between nucleotide sequences. Sequence similarity is calculated based on a reference sequence, which may be a subset of a larger sequence, such as a conserved motif, coding region, flanking region, etc. A reference sequence will usually be at least about 18 nt long, more usually at least about 30 nt long, and may extend to the complete
15 sequence that is being compared. Algorithms for sequence analysis are known in the art, such as BLAST, described in Altschul *et al.* (1990), *J. Mol. Biol.* **215**:403-10. Unless specified otherwise, all sequence analysis numbers provided herein are as determined with the BLAST program using default settings. The sequences provided herein are essential for recognizing *LGR4*, *LGR5* and *LGR7*-related and homologous proteins in database
20 searches.

LGR4, *LGR5* and *LGR7* NUCLEIC ACID COMPOSITIONS

- Nucleic acids encoding *LGR4*, *LGR5* and *LGR7* may be cDNA or genomic DNA or a fragment thereof. The terms "*LGR4* gene," "*LGR5* gene" and "*LGR7* gene" shall be
25 intended to mean the open reading frame encoding specific *LGR4*, *LGR5* and *LGR7* polypeptides, and *LGR4*, *LGR5* and *LGR7* introns, as well as adjacent 5' and 3' non-coding nucleotide sequences involved in the regulation of expression, up to about 20 kb beyond the coding region, but possibly further in either direction. The gene may be introduced into an appropriate vector for extra-chromosomal maintenance or for integration into a
30 host genome.

The term "cDNA" as used herein is intended to include all nucleic acids that share the arrangement of sequence elements found in native mature mRNA species, where sequence elements are exons and 3' and 5' non-coding regions. Normally mRNA species have contiguous exons, with the intervening introns, when present, removed by nuclear
5 RNA splicing, to create a continuous open reading frame encoding an LGR4, LGR5 and LGR7 protein.

A genomic sequence of interest comprises the nucleic acid present between the initiation codon and the stop codon, as defined in the listed sequences, including all of the introns that are normally present in a native chromosome. It may further include the 3'
10 and 5' untranslated regions found in the mature mRNA. It may further include specific transcriptional and translational regulatory sequences, such as promoters, enhancers, etc., including about 1 kb, but possibly more, of flanking genomic DNA at either the 5' or 3' end of the transcribed region. The genomic DNA may be isolated as a fragment of 100 kbp or smaller; and substantially free of flanking chromosomal sequence. The genomic
15 DNA flanking the coding region, either 3' or 5', or internal regulatory sequences as sometimes found in introns, contains sequences required for proper tissue and stage specific expression.

The sequence of the 5' flanking region may be utilized for promoter elements, including enhancer binding sites, that provide for developmental regulation in tissues
20 where *LGR4*, *LGR5* and/or *LGR7* is expressed. The tissue specific expression is useful for determining the pattern of expression, and for providing promoters that mimic the native pattern of expression. Naturally occurring polymorphisms in the promoter region are useful for determining natural variations in expression, particularly those that may be associated with disease.

Alternatively, mutations may be introduced into the promoter region to determine the effect of altering expression in experimentally defined systems. Methods for the identification of specific DNA motifs involved in the binding of transcriptional factors are known in the art, e.g. sequence similarity to known binding motifs, gel retardation studies, etc. For examples, see Blackwell *et al.* (1995), *Mol. Med.* 1:194-205; Mortlock *et al.*
30 (1996), *Genome Res.* 6:327-33; and Joulin and Richard-Foy (1995), *Eur. J. Biochem.* 232:620-626.

The regulatory sequences may be used to identify *cis* acting sequences required for transcriptional or translational regulation of *LGR4*, *LGR5* and/or *LGR7* expression, especially in different tissues or stages of development, and to identify *cis* acting sequences and *trans*-acting factors that regulate or mediate *LGR4*, *LGR* and/or *LGR7*

- 5 expression. Such transcription or translational control regions may be operably linked to an *LGR4*, *LGR5* or *LGR7* gene in order to promote expression of wild type or altered *LGR4*, *LGR5* or *LGR7* or other proteins of interest in cultured cells, or in embryonic, fetal or adult tissues, and for gene therapy.

- The nucleic acid compositions of the subject invention may encode all or a part of
10 the subject polypeptides. Double or single stranded fragments may be obtained of the DNA sequence by chemically synthesizing oligonucleotides in accordance with conventional methods, by restriction enzyme digestion, by PCR amplification, *etc.* For the most part, DNA fragments will be of at least 15 nt, usually at least 18 nt or 25 nt, and may be at least about 50 nt. Such small DNA fragments are useful as primers for PCR,
15 hybridization screening probes, *etc.* Larger DNA fragments, *i.e.* greater than 100 nt are useful for production of the encoded polypeptide. For use in amplification reactions, such as PCR, a pair of primers will be used. The exact composition of the primer sequences is not critical to the invention, but for most applications the primers will hybridize to the subject sequence under stringent conditions, as known in the art. It is preferable to choose
20 a pair of primers that will generate an amplification product of at least about 50 nt, preferably at least about 100 nt. Algorithms for the selection of primer sequences are generally known, and are available in commercial software packages. Amplification primers hybridize to complementary strands of DNA, and will prime towards each other.

- The *LGR4*, *LGR* and *LGR7* genes are isolated and obtained in substantial purity,
25 generally as other than an intact chromosome. Usually, the DNA will be obtained substantially free of other nucleic acid sequences that do not include an *LGR4*, *LGR5* or *LGR7* sequence or fragment thereof, generally being at least about 50%, usually at least about 90% pure and are typically "recombinant", *i.e.* flanked by one or more nucleotides with which it is not normally associated on a naturally occurring chromosome.

- 30 The DNA may also be used to identify expression of the gene in a biological specimen. The manner in which one probes cells for the presence of particular nucleotide

sequences, as genomic DNA or RNA, is well established in the literature and does not require elaboration here. DNA or mRNA is isolated from a cell sample. The mRNA may be amplified by RT-PCR, using reverse transcriptase to form a complementary DNA strand, followed by polymerase chain reaction amplification using primers specific for the subject DNA sequences. Alternatively, the mRNA sample is separated by gel electrophoresis, transferred to a suitable support, *e.g.* nitrocellulose, nylon, *etc.*, and then probed with a fragment of the subject DNA as a probe. Other techniques, such as oligonucleotide ligation assays, *in situ* hybridizations, and hybridization to DNA probes arrayed on a solid chip may also find use. Detection of mRNA hybridizing to the subject sequence is indicative of *LGR4*, *LGR5* and/or *LGR7* gene expression in the sample.

The sequence of an *LGR4*, *LGR5* or *LGR7* gene, including flanking promoter regions and coding regions, may be mutated in various ways known in the art to generate targeted changes in promoter strength, sequence of the encoded protein, *etc.* The DNA sequence or protein product of such a mutation will usually be substantially similar to the sequences provided herein, *i.e.* will differ by at least one nucleotide or amino acid, respectively, and may differ by at least two but not more than about ten nucleotides or amino acids. The sequence changes may be substitutions, insertions, deletions, or a combination thereof. Deletions may further include larger changes, such as deletions of a domain or exon. Other modifications of interest include epitope tagging, *e.g.* with the FLAG system, HA, *etc.* For studies of subcellular localization, fusion proteins with green fluorescent proteins (GFP) may be used.

Techniques for *in vitro* mutagenesis of cloned genes are known. Examples of protocols for site specific mutagenesis may be found in Gustin *et al.* (1993), *Biotechniques* 14:22; Barany (1985), *Gene* 37:111-23; Colicelli *et al.* (1985), *Mol. Gen. Genet.* 199:537-9; and Prentki *et al.* (1984), *Gene* 29:303-13. Methods for site specific mutagenesis can be found in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, CSH Press 1989, pp. 15.3-15.108; Weiner *et al.* (1993), *Gene* 126:35-41; Sayers *et al.* (1992), *Biotechniques* 13:592-6; Jones and Winistorfer (1992), *Biotechniques* 12:528-30; Barton *et al.* (1990), *Nucleic Acids Res* 18:7349-55; Marotti and Tomich (1989), *Gene Anal. Tech.* 6:67-70; and Zhu (1989), *Anal Biochem* 177:120-4. Such mutated genes may

be used to study structure-function relationships of LGR4, LGR5 and/or LGR7, or to alter properties of the protein that affect its function or regulation.

LGR4, LGR5 and LGR7 POLYPEPTIDES

- 5 Also provided by the subject invention are LGR4, LGR5 and LGR7 polypeptide compositions. The term polypeptide composition as used herein refers to both the full length proteins as well as portions or fragments thereof. Also included in this term are variations of the naturally occurring proteins, where such variations are homologous or substantially similar to the naturally occurring protein, be the naturally occurring protein
- 10 the human protein, mouse protein, or protein from some other species which naturally expresses an LGR4, LGR5 or LGR7 protein, usually a mammalian species. A candidate homologous protein is substantially similar to an LGR4, LGR5 or LGR7 protein of the subject invention, and therefore is an LGR4, LGR5 or LGR7 protein of the subject invention, if the candidate protein has a sequence that has at least about 80%, usually at
- 15 least about 90% and more usually at least about 98% sequence identity with an LGR4, LGR5 or LGR7 protein, as measured by BLAST, *supra*. In the following description of the subject invention, the term "LGR4, LGR5 or LGR7-protein" is used to refer not only to the human LGR4, LGR5 or LGR7 protein, but also to homologs thereof expressed in non-human species, e.g. murine, rat and other mammalian species.
- 20 The subject gene may be employed for producing all or portions of LGR4, LGR5 and LGR7 polypeptides. By "LGR4 polypeptide/protein", "LGR5 polypeptide/protein," and "LGR7 polypeptide/protein" is meant an amino acid sequence encoded by an open reading frame (ORF) of *LGR4*, *LGR5* and *LGR7* genes, including the full-length native polypeptide and fragments thereof, particularly biologically active fragments and/or
- 25 fragments corresponding to functional domains, e.g. extra-cellular regions; and including fusions of the subject polypeptides to other proteins or parts thereof, e.g. chimeric proteins. For expression, an expression cassette may be employed. The expression vector will provide a transcriptional and translational initiation region, which may be inducible or constitutive, where the coding region is operably linked under the transcriptional control
- 30 of the transcriptional initiation region, and a transcriptional and translational termination

region. These control regions may be native to an *LGR4*, *LGR5* or *LGR7* gene, or may be derived from exogenous sources.

Expression vectors generally have convenient restriction sites located near the promoter sequence to provide for the insertion of nucleic acid sequences encoding heterologous proteins. A selectable marker operative in the expression host may be present. Expression vectors may be used for the production of fusion proteins, where the exogenous fusion peptide provides additional functionality, i.e. increased protein synthesis, stability, reactivity with defined antisera, an enzyme marker, e.g. β -galactosidase, etc.

Expression cassettes may be prepared comprising a transcription initiation region, the gene or fragment thereof, and a transcriptional termination region. Of particular interest is the use of sequences that allow for the expression of functional epitopes or domains, usually at least about 8 amino acids in length, more usually at least about 15 amino acids in length, to about 25 amino acids, and up to the complete open reading frame of the gene. After introduction of the DNA, the cells containing the construct may be selected by means of a selectable marker, the cells expanded and then used for expression.

LGR4, *LGR5* or *LGR7* polypeptides may be expressed in prokaryotes or eukaryotes in accordance with conventional ways, depending upon the purpose for expression. For large scale production of the protein, a unicellular organism, such as *E. coli*, *B. subtilis*, *S. cerevisiae*, insect cells in combination with baculovirus vectors, or cells of a higher organism such as vertebrates, particularly mammals, e.g. COS 7 cells, may be used as the expression host cells. In some situations, it is desirable to express the *LGR4*, *LGR5* or *LGR7* gene in eukaryotic cells, where the *LGR4*, *LGR5* or *LGR7* protein will benefit from native folding and post-translational modifications. Small peptides can also be synthesized in the laboratory. Polypeptides that are subsets of the complete *LGR4*, *LGR5* or *LGR7* sequence may be used to identify and investigate parts of the protein important for function or to raise antibodies directed against these regions.

For production of the extracellular domain of the *LGR4*, *LGR5* or *LGR7* receptor, the anchored receptor approach as described in Osuga et al, Mol. Endocrinol. (1997) 11: 1659-1668 may be employed. Likewise, the chimeric receptor approach described in Kudo et al, J Biol. Chem. (1996) 271: 22470-22478 may be used.

Such peptides find use in the identification of endogenous ligands and in drug screening for agonists and antagonists using methods described in Osuga, supra. Solubilized extracellular domains find use as therapeutic agents, e.g. in the neutralization of the action of endogenous ligands.

- 5 With the availability of the protein or fragments thereof in large amounts, by employing an expression host, the protein may be isolated and purified in accordance with conventional ways. A lysate may be prepared of the expression host and the lysate purified using HPLC, exclusion chromatography, gel electrophoresis, affinity chromatography, or other purification technique. The purified protein will generally be at
10 least about 80% pure, preferably at least about 90% pure, and may be up to and including 100% pure. Pure is intended to mean free of other proteins, as well as cellular debris.

- The expressed LGR4, LGR5 and LGR7 polypeptides are useful for the production of antibodies, where short fragments provide for antibodies specific for the particular polypeptide, and larger fragments or the entire protein allow for the production of
15 antibodies over the surface of the polypeptide. Antibodies may be raised to the wild-type or variant forms of LGR4, LGR5 or LGR7. Antibodies may be raised to isolated peptides corresponding to these domains, or to the native protein.

- Antibodies are prepared in accordance with conventional ways, where the expressed polypeptide or protein is used as an immunogen, by itself or conjugated to
20 known immunogenic carriers, e.g. KLH, pre-S HBsAg, other viral or eukaryotic proteins, or the like. Various adjuvants may be employed, with a series of injections, as appropriate. Both polyclonal and monoclonal antibodies may be produced. For monoclonal antibodies, after one or more booster injections, the spleen is isolated, the lymphocytes immortalized by cell fusion, and then screened for high affinity antibody
25 binding. The immortalized cells, i.e. hybridomas, producing the desired antibodies may then be expanded. For further description, see Monoclonal Antibodies: A Laboratory Manual, Harlow and Lane eds., Cold Spring Harbor Laboratories, Cold Spring Harbor, New York, 1988. If desired, the mRNA encoding the heavy and light chains may be isolated and mutagenized by cloning in *E. coli*, and the heavy and light chains mixed to
30 further enhance the affinity of the antibody. Alternatives to *in vivo* immunization as a

method of raising antibodies include binding to phage "display" libraries, usually in conjunction with *in vitro* affinity maturation.

DIAGNOSTIC USES

5 The subject nucleic acid and/or polypeptide compositions may be used to analyze a patient sample for the presence of polymorphisms associated with a disease state or genetic predisposition to a disease state. Biochemical studies may be performed to determine whether a sequence polymorphism in an *LGR4*, *LGR* or *LGR7* coding region or control regions is associated with disease. Disease associated polymorphisms may include
10 deletion or truncation of the gene, mutations that alter expression level, that affect the activity of the protein, and the like.

 Changes in the promoter or enhancer sequence that may affect expression levels of *LGR4*, *LGR5* or *LGR7* can be compared to expression levels of the normal allele by various methods known in the art. Methods for determining promoter or enhancer
15 strength include quantitation of the expressed natural protein; insertion of the variant control element into a vector with a reporter gene such as β -galactosidase, luciferase, chloramphenicol acetyltransferase, *etc.*, that provides for convenient quantitation; and the like.

 A number of methods are available for analyzing nucleic acids for the presence of
20 a specific sequence, *e.g.* a disease associated polymorphism. Where large amounts of DNA are available, genomic DNA is used directly. Alternatively, the region of interest is cloned into a suitable vector and grown in sufficient quantity for analysis. Cells that express *LGR4*, *LGR5* or *LGR7* may be used as a source of mRNA, which may be assayed directly or reverse transcribed into cDNA for analysis. The nucleic acid may be amplified
25 by conventional techniques, such as the polymerase chain reaction (PCR), to provide sufficient amounts for analysis. The use of the polymerase chain reaction is described in Saiki, *et al.* (1985), *Science* **239**:487, and a review of techniques may be found in Sambrook, *et al.* Molecular Cloning: A Laboratory Manual, CSH Press 1989, pp.14.2–14.33. Alternatively, various methods are known in the art that utilize oligonucleotide
30 ligation as a means of detecting polymorphisms, for examples see Riley *et al.* (1990),

Nucl. Acids Res. **18**:2887-2890; and Delahunty *et al.* (1996), *Am. J. Hum. Genet.* **58**:1239-1246.

A detectable label may be included in an amplification reaction. Suitable labels include fluorochromes, *e.g.* fluorescein isothiocyanate (FITC), rhodamine, Texas Red, phycoerythrin, allophycocyanin, 6-carboxyfluorescein (6-FAM), 2',7'-dimethoxy-4',5'-dichloro-6-carboxyfluorescein (JOE), 6-carboxy-X-rhodamine (ROX), 6-carboxy-2',4',7',4',7'-hexachlorofluorescein (HEX), 5-carboxyfluorescein (5-FAM) or N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA), radioactive labels, *e.g.* ³²P, ³⁵S, ³H; *etc.* The label may be a two stage system, where the amplified DNA is conjugated to biotin, haptens, *etc.* having a high affinity binding partner, *e.g.* avidin, specific antibodies, *etc.*, where the binding partner is conjugated to a detectable label. The label may be conjugated to one or both of the primers. Alternatively, the pool of nucleotides used in the amplification is labeled, so as to incorporate the label into the amplification product.

The sample nucleic acid, *e.g.* amplified or cloned fragment, is analyzed by one of a number of methods known in the art. The nucleic acid may be sequenced by dideoxy or other methods, and the sequence of bases compared to a wild-type *LGR4*, *LGR5* or *LGR7* sequence. Hybridization with the variant sequence may also be used to determine its presence, by Southern blots, dot blots, *etc.* The hybridization pattern of a control and variant sequence to an array of oligonucleotide probes immobilized on a solid support, as described in US 5,445,934, or in WO 95/35505 (the disclosures of which are herein incorporated by reference), may also be used as a means of detecting the presence of variant sequences. Single strand conformational polymorphism (SSCP) analysis, denaturing gradient gel electrophoresis (DGGE), and heteroduplex analysis in gel matrices are used to detect conformational changes created by DNA sequence variation as alterations in electrophoretic mobility. Alternatively, where a polymorphism creates or destroys a recognition site for a restriction endonuclease, the sample is digested with that endonuclease, and the products size fractionated to determine whether the fragment was digested. Fractionation is performed by gel or capillary electrophoresis, particularly acrylamide or agarose gels.

Screening for mutations in *LGR4*, *LGR5* or *LGR7* may be based on the functional or antigenic characteristics of the protein. Protein truncation assays are useful in detecting

deletions that may affect the biological activity of the protein. Various immunoassays designed to detect polymorphisms in LGR4, LGR5 or LGR7 proteins may be used in screening. Where many diverse genetic mutations lead to a particular disease phenotype, functional protein assays have proven to be effective screening tools. The activity of the encoded LGR4, LGR5 or LGR7 protein may be determined by comparison with the wild-type protein.

Antibodies specific for LGR4, LGR5 or LGR7 proteins may be used in staining or in immunoassays. Samples, as used herein, include biological fluids such as semen, blood, cerebrospinal fluid, tears, saliva, lymph, dialysis fluid and the like; organ or tissue culture derived fluids; and fluids extracted from physiological tissues. Also included in the term are derivatives and fractions of such fluids. The cells may be dissociated, in the case of solid tissues, or tissue sections may be analyzed. Alternatively a lysate of the cells may be prepared.

Diagnosis may be performed by a number of methods to determine the absence or presence or altered amounts of normal or abnormal LGR4, LGR5 or LGR7 in patient cells. For example, detection may utilize staining of cells or histological sections, performed in accordance with conventional methods. Cells are permeabilized to stain cytoplasmic molecules. The antibodies of interest are added to the cell sample, and incubated for a period of time sufficient to allow binding to the epitope, usually at least about 10 minutes. The antibody may be labeled with radioisotopes, enzymes, fluorescers, chemiluminescers, or other labels for direct detection. Alternatively, a second stage antibody or reagent is used to amplify the signal. Such reagents are well known in the art. For example, the primary antibody may be conjugated to biotin, with horseradish peroxidase-conjugated avidin added as a second stage reagent. Alternatively, the secondary antibody conjugated to a fluorescent compound, *e.g.* fluorescein, rhodamine, Texas red, *etc.* Final detection uses a substrate that undergoes a color change in the presence of the peroxidase. The absence or presence of antibody binding may be determined by various methods, including flow cytometry of dissociated cells, microscopy, radiography, scintillation counting, *etc.*

Diagnostic screening may also be performed for polymorphisms that are genetically linked to a disease predisposition, particularly through the use of microsatellite

- markers or single nucleotide polymorphisms. Frequently the microsatellite polymorphism itself is not phenotypically expressed, but is linked to sequences that result in a disease predisposition. However, in some cases the microsatellite sequence itself may affect gene expression. Microsatellite linkage analysis may be performed alone, or in combination
- 5 with direct detection of polymorphisms, as described above. The use of microsatellite markers for genotyping is well documented. For examples, see Mansfield *et al.* (1994), *Genomics* 24:225-233; Ziegler *et al.* (1992), *Genomics* 14:1026-1031; Dib *et al.*, *supra*.

MODULATION OF *LGR4*, *LGR5* and *LGR7* GENE EXPRESSION

- 10 The *LGR4*, *LGR5* or *LGR7* genes, gene fragments, or the *LGR4*, *LGR5* or *LGR7* protein or protein fragments, are useful in gene therapy to treat disorders associated with *LGR4*, *LGR5* or *LGR7* defects. Expression vectors may be used to introduce the *LGR4*, *LGR5* or *LGR7* gene into a cell. Such vectors generally have convenient restriction sites located near the promoter sequence to provide for the insertion of nucleic acid sequences.
- 15 Transcription cassettes may be prepared comprising a transcription initiation region, the target gene or fragment thereof, and a transcriptional termination region. The transcription cassettes may be introduced into a variety of vectors, *e.g.* plasmid; retrovirus, *e.g.* lentivirus; adenovirus; and the like, where the vectors are able to transiently or stably be maintained in the cells, usually for a period of at least about one day, more usually for a
- 20 period of at least about several days to several weeks.

- The gene or *LGR4*, *LGR5* or *LGR7* protein may be introduced into tissues or host cells by any number of routes, including viral infection, microinjection, or fusion of vesicles. Jet injection may also be used for intramuscular administration, as described by Furth *et al.* (1992), *Anal Biochem* 205:365-368. The DNA may be coated onto gold
- 25 microparticles, and delivered intradermally by a particle bombardment device, or "gene gun" as described in the literature (see, for example, Tang *et al.* (1992), *Nature* 356:152-154), where gold microprojectiles are coated with the *LGR4*, *LGR5* or *LGR7* DNA, then bombarded into skin cells.

- Antisense molecules can be used to down-regulate expression of *LGR4*, *LGR5*, or
- 30 *LGR7* in cells. The anti-sense reagent may be antisense oligonucleotides (ODN), particularly synthetic ODN having chemical modifications from native nucleic acids, or

nucleic acid constructs that express such anti-sense molecules as RNA. The antisense sequence is complementary to the mRNA of the targeted gene, and inhibits expression of the targeted gene products. Antisense molecules inhibit gene expression through various mechanisms, e.g. by reducing the amount of mRNA available for translation, through
5 activation of RNase H, or steric hindrance. One or a combination of antisense molecules may be administered, where a combination may comprise multiple different sequences.

Antisense molecules may be produced by expression of all or a part of the target gene sequence in an appropriate vector, where the transcriptional initiation is oriented such that an antisense strand is produced as an RNA molecule. Alternatively, the
10 antisense molecule is a synthetic oligonucleotide. Antisense oligonucleotides will generally be at least about 7, usually at least about 12, more usually at least about 20 nucleotides in length, and not more than about 500, usually not more than about 50, more usually not more than about 35 nucleotides in length, where the length is governed by efficiency of inhibition, specificity, including absence of cross-reactivity, and the like. It
15 has been found that short oligonucleotides, of from 7 to 8 bases in length, can be strong and selective inhibitors of gene expression (see Wagner *et al.* (1996), *Nature Biotechnol.* **14**:840-844).

A specific region or regions of the endogenous sense strand mRNA sequence is chosen to be complemented by the antisense sequence. Selection of a specific sequence
20 for the oligonucleotide may use an empirical method, where several candidate sequences are assayed for inhibition of expression of the target gene in an *in vitro* or animal model. A combination of sequences may also be used, where several regions of the mRNA sequence are selected for antisense complementation.

Antisense oligonucleotides may be chemically synthesized by methods known in
25 the art (see Wagner *et al.* (1993), *supra*, and Milligan *et al.*, *supra*.) Preferred oligonucleotides are chemically modified from the native phosphodiester structure, in order to increase their intracellular stability and binding affinity. A number of such modifications have been described in the literature, which alter the chemistry of the backbone, sugars or heterocyclic bases.

30 Among useful changes in the backbone chemistry are phosphorothioates; phosphorodithioates, where both of the non-bridging oxygens are substituted with sulfur;

phosphoroamidites; alkyl phosphotriesters and boranophosphates. Achiral phosphate derivatives include 3'-O'-5'-S-phosphorothioate, 3'-S-5'-O-phosphorothioate, 3'-CH₂-5'-O-phosphonate and 3'-NH-5'-O-phosphoroamidate. Peptide nucleic acids replace the entire ribose phosphodiester backbone with a peptide linkage. Sugar modifications are also used to enhance stability and affinity. The α -anomer of deoxyribose may be used, where the base is inverted with respect to the natural β -anomer. The 2'-OH of the ribose sugar may be altered to form 2'-O-methyl or 2'-O-allyl sugars, which provides resistance to degradation without comprising affinity. Modification of the heterocyclic bases must maintain proper base pairing. Some useful substitutions include deoxyuridine for deoxythymidine; 5-methyl-2'-deoxycytidine and 5-bromo-2'-deoxycytidine for deoxycytidine. 5-propynyl-2'-deoxyuridine and 5-propynyl-2'-deoxycytidine have been shown to increase affinity and biological activity when substituted for deoxythymidine and deoxycytidine, respectively.

As an alternative to anti-sense inhibitors, catalytic nucleic acid compounds, *e.g.* ribozymes, anti-sense conjugates, *etc.* may be used to inhibit gene expression. Ribozymes may be synthesized *in vitro* and administered to the patient, or may be encoded on an expression vector, from which the ribozyme is synthesized in the targeted cell (for example, see International patent application WO 9523225, and Beigelman *et al.* (1995), *Nucl. Acids Res.* **23**:4434-42). Examples of oligonucleotides with catalytic activity are described in WO 9506764. Conjugates of anti-sense ODN with a metal complex, *e.g.* terpyridylCu(II), capable of mediating mRNA hydrolysis are described in Bashkin *et al.* (1995), *Appl. Biochem. Biotechnol.* **54**:43-56.

GENETICALLY ALTERED CELL OR ANIMAL MODELS FOR LGR4, LGR5 AND LGR7 FUNCTION

The subject nucleic acids can be used to generate transgenic, non-human animals or site specific gene modifications in cell lines. Transgenic animals may be made through homologous recombination, where the normal *LGR4*, *LGR5* or *LGR7* locus is altered. Alternatively, a nucleic acid construct is randomly integrated into the genome. Vectors for stable integration include plasmids, retroviruses and other animal viruses, YACs, and the like.

The modified cells or animals are useful in the study of *LGR4*, *LGR5* and/or *LGR7* function and regulation. For example, a series of small deletions and/or substitutions may be made in the host's native *LGR4*, *LGR5* or *LGR7* gene to determine the role of different exons. Of interest is the use of *LGR4*, *LGR5* or *LGR7* to construct transgenic animal models for disease states. Specific constructs of interest include anti-sense *LGR4*, *LGR5* or *LGR7*, which will block *LGR4*, *LGR5* or *LGR7* expression, expression of dominant negative *LGR4*, *LGR5* or *LGR7* mutations, and over-expression of *LGR4*, *LGR5* or *LGR7* genes. Where an *LGR4*, *LGR5* or *LGR7* sequence is introduced, the introduced sequence may be either a complete or partial sequence of an *LGR4*, *LGR5* or *LGR7* gene native to the host, or may be a complete or partial *LGR4*, *LGR5* or *LGR7* sequence that is exogenous to the host animal. *e.g.*, a human *LGR4*, *LGR5* or *LGR7* sequence. A detectable marker, such as *lac Z* may be introduced into the *LGR4*, *LGR5* or *LGR7* locus, where upregulation of *LGR4*, *LGR5* or *LGR7* expression will result in an easily detected change in phenotype.

One may also provide for expression of the *LGR4*, *LGR5* or *LGR7* gene or variants thereof in cells or tissues where it is not normally expressed, at levels not normally present in such cells or tissues, or at abnormal times of development. By providing expression of *LGR4*, *LGR5* or *LGR7* protein in cells in which it is not normally produced, one can induce changes in cell behavior, *e.g.* through *LGR4*, *LGR5* or *LGR7* mediated activity.

DNA constructs for homologous recombination will comprise at least a portion of the *LGR4*, *LGR5* or *LGR7* gene, which may or may not be native to the species of the host animal, wherein the gene has the desired genetic modification(s), and includes regions of homology to the target locus. DNA constructs for random integration need not include regions of homology to mediate recombination. Conveniently, markers for positive and negative selection are included. Methods for generating cells having targeted gene modifications through homologous recombination are known in the art. For various techniques for transfecting mammalian cells, see Keown *et al.* (1990), *Meth. Enzymol.* **185**:527-537.

For embryonic stem (ES) cells, an ES cell line may be employed, or embryonic cells may be obtained freshly from a host, *e.g.* mouse, rat, guinea pig, *etc.* Such cells are grown on an appropriate fibroblast-feeder layer or grown in the presence of leukemia

inhibiting factor (LIF). When ES or embryonic cells have been transformed, they may be used to produce transgenic animals. After transformation, the cells are plated onto a feeder layer in an appropriate medium. Cells containing the construct may be detected by employing a selective medium. After sufficient time for colonies to grow, they are picked and analyzed for the occurrence of homologous recombination or integration of the construct. Those colonies that are positive may then be used for embryo manipulation and blastocyst injection. Blastocysts are obtained from 4 to 6 week old superovulated females. The ES cells are trypsinized, and the modified cells are injected into the blastocoel of the blastocyst. After injection, the blastocysts are returned to each uterine horn of pseudopregnant females. Females are then allowed to go to term and the resulting offspring screened for the construct. By providing for a different phenotype of the blastocyst and the genetically modified cells, chimeric progeny can be readily detected.

The chimeric animals are screened for the presence of the modified gene and males and females having the modification are mated to produce homozygous progeny. If the gene alterations cause lethality at some point in development, tissues or organs can be maintained as allogeneic or congenic grafts or transplants, or in *in vitro* culture. The transgenic animals may be any non-human mammal, such as laboratory animals, domestic animals, etc. The transgenic animals may be used in functional studies, drug screening, etc., e.g. to determine the effect of a candidate drug on *LGR4*, *LGR5* or *LGR7* or related gene activation etc.

IN VITRO MODELS FOR *LGR4*, *LGR5* or *LGR7* FUNCTION

The availability of a number of components in the G-protein coupled receptor family, as previously described, allows *in vitro* reconstruction of the processes or systems in which members of this family operate. Two or more of the components, such as the isolated receptor and a potential ligand therefore, may be combined *in vitro*, and the behavior assessed in terms of activation of transcription of specific target sequences; modification of protein components, e.g. proteolytic processing, phosphorylation, methylation, etc.; ability of different protein components to bind to each other. The components may be modified by sequence deletion, substitution, etc. to determine the functional role of specific domains.

Drug screening may be performed using an *in vitro* model, a genetically altered cell or animal, purified LGR4, LGR5 or LGR7 protein, as well as fragments or portions thereof, e.g. solubilized extra-cellular domain or chimeric receptor proteins comprising the LGR4, LGR5 or LGR7 extra-cellular domain. One can identify ligands or substrates that bind to and modulate the action of LGR4, LGR5 or LGR7. Areas of investigation include the development of agents that beneficially counter abnormalities related to LGR4, LGR5 or LGR7 and the use of such agents in the therapy.

Drug screening identifies agents that modulate the activity of LGR4, LGR5 or LGR7 function in abnormal cells. Of particular interest are screening assays for agents that have a low toxicity for human cells. A wide variety of assays may be used for this purpose, including labeled *in vitro* protein-protein binding assays, electrophoretic mobility shift assays, immunoassays for protein binding, and the like. The purified protein may also be used for determination of three-dimensional crystal structure, which can be used for modeling intermolecular interactions, such as GTP binding, etc.

The term "agent" as used herein describes any molecule, e.g. protein or pharmaceutical, with the capability of altering or mimicking the physiological function of LGR4, LGR5 or LGR7. Generally a plurality of assay mixtures are run in parallel with different agent concentrations to obtain a differential response to the various concentrations. Typically, one of these concentrations serves as a negative control, i.e. at zero concentration or below the level of detection.

In some embodiments, candidate agents encompass numerous chemical classes, though typically they are organic molecules, preferably small organic compounds having a molecular weight of more than 50 and less than about 2,500 daltons. Candidate agents comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The candidate agents often comprise cyclical carbon or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups. Candidate agents are also found among biomolecules including peptides, saccharides, fatty acids, steroids, purines, pyrimidines, derivatives, structural analogs or combinations thereof.

Candidate agents are obtained from a wide variety of sources including libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides and oligopeptides. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily produced. Additionally, natural or synthetically produced libraries and compounds are readily modified through conventional chemical, physical and biochemical means, and may be used to produce combinatorial libraries. Known pharmacological agents may be subjected to directed or random chemical modifications, such as acylation, alkylation, esterification, amidification, *etc.* to produce structural analogs.

Of particular interest in certain embodiments are peptidic agents based on LGR4, LGR5 or LGR7, e.g. solubilized extra-cellular domain or chimeric receptor proteins comprising the LGR4, LGR5 or LGR7 extra-cellular domain, where such agents neutralize the activity of endogenous LGR4, LGR5 or LGR7 ligands, e.g. hormones.

Where the screening assay is a binding assay, one or more of the molecules may be joined to a label, where the label can directly or indirectly provide a detectable signal. Various labels include radioisotopes, fluorescers, chemiluminescers, enzymes, specific binding molecules, particles, *e.g.* magnetic particles, and the like. Specific binding molecules include pairs, such as biotin and streptavidin, digoxin and antidigoxin *etc.* For the specific binding members, the complementary member would normally be labeled with a molecule that provides for detection, in accordance with known procedures.

A variety of other reagents may be included in the screening assay. These include reagents like salts, neutral proteins, *e.g.* albumin, detergents, *etc.*, that are used to facilitate optimal protein-protein binding and/or reduce non-specific or background interactions. Reagents that improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, anti-microbial agents, *etc.*, may be used. The mixture of components are added in any order that provides for the requisite binding. Incubations are performed at any suitable temperature, typically between 4 and 40°C. Incubation periods are selected for optimum activity, but may also be optimized to facilitate rapid high-throughput screening. Typically between 0.1 and 1 hours will be sufficient.

Other assays of interest detect agents that mimic LGR4, LGR5 or LGR7 function. For example, an expression construct comprising an *LGR4*, *LGR5* or *LGR7* gene may be introduced into a cell line under conditions that allow expression. The level of LGR4, LGR5 or LGR7 activity is determined by a functional assay, as previously described. In one screening assay, the ability of candidate agents to inhibit or enhance LGR4, LGR5 or LGR7 function is determined. Alternatively, candidate agents are added to a cell that lacks functional LGR4, LGR5 or LGR7, and screened for the ability to reproduce LGR4, LGR5 or LGR7 activity in a functional assay.

The compounds having the desired pharmacological activity may be administered in a physiologically acceptable carrier to a host for treatment, *etc.* The compounds may also be used to enhance *LGR4*, *LGR5* or *LGR7* function. The inhibitory agents may be administered in a variety of ways, orally, topically, parenterally *e.g.* subcutaneously, intraperitoneally, by viral infection, intravascularly, *etc.* Topical treatments are of particular interest. Depending upon the manner of introduction, the compounds may be formulated in a variety of ways. The concentration of therapeutically active compound in the formulation may vary from about 0.1-100 wt.%. 15

The pharmaceutical compositions can be prepared in various forms, such as granules, tablets, pills, suppositories, capsules, suspensions, salves, lotions and the like. Pharmaceutical grade organic or inorganic carriers and/or diluents suitable for oral and topical use can be used to make up compositions containing the therapeutically-active compounds. Diluents known to the art include aqueous media, vegetable and animal oils and fats. Stabilizing agents, wetting and emulsifying agents, salts for varying the osmotic pressure or buffers for securing an adequate pH value, and skin penetration enhancers can be used as auxiliary agents. 25

EXPERIMENTAL

The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the subject invention, and are not intended to limit the scope of what is regarded as the invention. Efforts have been made to ensure accuracy with respect to the numbers used (*e.g.* amounts, temperature, concentrations, *etc.*) but some experimental errors and deviations 30

should be allowed for. Unless otherwise indicated, parts are parts by weight, molecular weight is average molecular weight, temperature is in degrees centigrade; and pressure is at or near atmospheric.

5 **Example 1. Identification of LGR4 and LGR5**

Human sequences related to the sea anemone and *Drosophila* glycoprotein hormone receptors were identified from the expression sequence tag database (dbEST) at the National Center for Biotechnology Information by using the BLAST server with the BLOSUM62 protein comparison matrix (Altschul SF *et al*, Nucleic Acids Res (1997) 10 25:3389-3402). Human ESTs showing high homology to two non-overlapping regions of the gonadotropin receptors were identified. Clones AA312798 and AA298810 were found to encode transmembrane four to five of the putative receptor LGR4 whereas AA460529 and AA424098 encode transmembrane two to three of the putative receptor LGR5. Using these ESTs to further search the GenBank EST division database, 15 overlapping EST sequences were aligned to obtain the longest open reading frame (ORF) for these receptors.

Based on the longest human ORF, specific primers were designed for PCR amplification of LGR4 and LGR5 cDNA fragments from rat ovary and human placenta, respectively. After hybridization with labeled EST clones and confirmation of DNA 20 sequences by dideoxy DNA sequencing, specific receptor fragments isolated were used to design primers to prepare sub-cDNA libraries enriched with specific receptor cDNAs. For 5' extension, reverse transcription was performed using rat ovarian and human placenta mRNA preparations and receptor-specific primers. Following second strand synthesis, the enriched cDNA pool was tailed at 5'-ends with specific adaptor sequences to allow further 25 PCR amplification. For 3' extension, rat ovarian or human placenta mRNAs were reversed transcribed using oligo-dT, followed by second strand synthesis using receptor-specific primers and adaptor tailing. These mini-libraries were further used as templates for PCR amplification of upstream or downstream cDNAs specific for each receptor using internal primers. PCR products with a strong hybridization signal to each receptor cDNA 30 fragment were subcloned into the pUC18 or pCDNA3 vectors. After screening of these sublibraries based on colony hybridization using specific receptor probes, clones with 5'-

or 3'-sequences of the putative receptors were identified and isolated for DNA sequencing. As needed, the procedure was repeated up to three times to generate cDNAs encoding the complete ORF of each putative receptor for sequence analysis and for the expression of receptor proteins in eukaryotic cells. The entire coding sequences of each gene were also amplified with specific primers flanking the entire ORF in independent experiments. At least three independent PCR clones were sequenced to verify the authenticity of coding sequences. The nucleotide sequence of LGR4, as well as the amino acid sequence of the product encoded by the ORF thereof, is provided in Fig. 1. The nucleotide sequence of LGR5, as well as the amino acid sequence of the product encoded by the ORF thereof, is provided in Fig. 2.

Example 2. Comparison of deduced amino acid sequence of LGR4 and 5 cDNAs and those encoding FSH and LH receptors.

Sequence alignment of LGR4 and LGR5 with known human glycoprotein hormone receptors was performed and the results are shown in Fig. 6. Shaded residues are identical in at least two of the four receptor proteins shown.

Example 3. Expression pattern of LGR4 and 5 mRNA transcripts in different tissues.

For northern blot analysis, poly (A)+-selected RNA from different human tissues was hybridized with a ³²P-labeled cDNA probes. After washing, the blots were exposed to X-ray films at -70°C for five days. Subsequent hybridization with a beta-actin cDNA probe was performed to estimate nucleic acid loading (8 h exposure). LGR4 was shown to be expressed in placenta, ovary, testis, adrenal, spinal cord, thyroid, stomach, trachea, heart, pancreas, kidney, prostate and spleen while LGR5 was shown to be expressed in the skeletal muscle, placenta, spinal cord, brain, adrenal, colon, stomach, ovary and bone marrow.

Example 4. Chromosomal localization of LGR4 and 5 in human.

Using genomic fragments of LGR4 (>100 Kb) and LGR5 (>100 Kb) as probes, chromosomal localization of these genes were detected using the FISH method to banded DNA in chromosomal 5q34-35.1 and 12q15, respectively.

Example 5. Identification of LGR7.

Analysis of EST databases has revealed a novel LGR closely related to a G protein-coupled receptor from pond snail (*Lymnaea stagnalis*, accession no. 481946). Because the snail G-protein coupled receptor shared the leucine-rich repeat ectodomain and seven transmembrane region characteristics of mammalian LGRs, the novel EST sequence could encode either a homologue of snail receptor or a novel mammalian LGR. For the isolation of LGR7 cDNA, a Clontech Marathon-ready testis cDNA pool was used as the template for 5' and 3' RACE with adapter and gene-specific primers. Sequence analysis of the RACE products showed that LGR7 gene encode at least two splicing variants differ at the N-terminus. The nucleotide sequence of the long variant, as well as the amino acid sequence of the product encoded by the ORF thereof, is provided in Fig. 3; while the nucleotide sequence of the short variant, as well as the amino acid sequence of the ORF thereof, is provided in Fig. 4. Both variants contain a classical C-terminal 7-transmembrane region and a leucine-rich repeat ectodomain flanked by cysteine rich regions found in other mammalian LGRs. The long form LGR7 contains extra 35 amino acids in the N-terminal cysteine rich region as compared to the short form LGR7. Of interest, analysis of the LGR7 ORF from either variant showed that its tertiary structure resembles that of mammalian LGRs instead of the snail receptor, which shares the greatest identity in the transmembrane region. These findings suggest that LGR7 and snail receptor diverged early during evolution and LGR7 perhaps adopted new function in higher organisms.

Based on the LGR7 cDNA sequence, we further identified a human genomic DNA fragment (AQ053279) in the genomic survey sequence division of GenBank that contains part of the LGR7 gene. The authenticity of this genomic clone was confirmed by Southern blot hybridization and the genomic clone was used as the probe to identify the chromosomal localization for LGR7 gene.

It is evident from the above discussion and results that three novel mammalian G-protein coupled receptors, as well as a nucleic acids encoding the same, are provided by the subject invention. The inventions described above find use in a variety of applications, including research and therapeutic applications.

5

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference. The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such a disclosure by virtue of prior invention.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

WHAT IS CLAIMED IS:

1. An isolated nucleic acid encoding a mammalian protein selected from the group consisting of LGR4, LGR5 or LGR7.
- 5
2. An isolated nucleic acid according to Claim 1, wherein said mammalian protein has the amino acid sequence of SEQ ID NO:2, SEQ ID NO:04, SEQ ID NO:06 or SEQ ID NO:08.
- 10
3. An isolated nucleic acid according to Claim 1, wherein said mammalian protein has an amino acid sequence that is substantially identical to the amino acid sequence of SEQ ID NO:2, SEQ ID NO:04, SEQ ID NO:06 or SEQ ID NO:08.
- 15
4. An isolated nucleic acid according to Claim 1, wherein the nucleotide sequence of said nucleic acid has the sequence selected from the group consisting of: (a) SEQ ID NO:1 or the complementary sequence thereof; (b) SEQ ID NO:03 or the complementary sequence thereof; (c) SEQ ID NO:05 or the complementary sequence thereof; and (d) SEQ ID NO:07 or the complementary sequence thereof.
- 20
5. An isolated nucleic acid comprising at least 18 contiguous nucleotides of the sequence selected from the group consisting of: (a) SEQ ID NO:1 or the complementary sequence thereof; (b) SEQ ID NO:03 or the complementary sequence thereof; (c) SEQ ID NO:05 or the complementary sequence thereof; and (d) SEQ ID NO:07 or the complementary sequence thereof.
- 25
6. An isolated nucleic acid comprising at least 50 contiguous nucleotides of the sequence selected from the group consisting of: (a) SEQ ID NO:1 or the complementary sequence thereof; (b) SEQ ID NO:03 or the complementary sequence thereof; (c) SEQ ID NO:05 or the complementary sequence thereof; and (d) SEQ ID NO:07 or the complementary sequence thereof.
- 30

7. An isolated nucleic acid that hybridizes under stringent conditions to a nucleic acid having the nucleotide sequence selected from the group consisting of: (a) SEQ ID NO:1 or the complementary sequence thereof; (b) SEQ ID NO:03 or the
5 complementary sequence thereof; (c) SEQ ID NO:05 or the complementary sequence thereof; and (d) SEQ ID NO:07 or the complementary sequence thereof.

8. An expression cassette comprising a transcriptional initiation region functional in an expression host, a nucleic acid having a sequence of the isolated nucleic
10 acid according to Claim 1 under the transcriptional regulation of said transcriptional initiation region, and a transcriptional termination region functional in said expression host.

9. A cell comprising an expression cassette according to Claim 8 as part of an
15 extrachromosomal element or integrated into the genome of a host cell as a result of introduction of said expression cassette into said host cell, and the cellular progeny of said host cell.

10. A method for producing a mammalian protein selected from the group
20 consisting of LGR4, LGR5 and LGR7, said method comprising:
growing a cell according to Claim 9, whereby said mammalian protein is expressed; and
isolating said protein substantially free of other proteins.

25 11. A purified polypeptide composition comprising at least 50 weight % of the protein present as a mammalian protein selected from the group consisting of LGR4, LGR5 and LGR7, or a fragment thereof.

12. An antibody binding specifically to a mammalian protein selected from the
30 group consisting of LGR4, LGR5 and LGR7.

13. The antibody of Claim 12, wherein said antibody is a monoclonal antibody.

14. A non-human transgenic animal model for *LGR4*, *LGR5* or *LGR7* gene function, wherein said transgenic animal comprises an introduced alteration in an *LGR4*,
5 *LGR5* or *LGR7* gene.

15. The animal model of claim 14, wherein said animal is heterozygous for said introduced alteration.

10 16. The animal model of claim 14, wherein said animal is homozygous for said introduced alteration.

17. The animal model of claim 14, wherein said introduced alteration is a knockout of endogenous *LGR4*, *LGR5* or *LGR7* gene expression.

15 18. A method of screening a sample for the presence of a ligand for a receptor selected from the group consisting of *LGR4*, *LGR5* and *LGR7*, said method comprising:
contacting said sample with a receptor selected from the group consisting of *LGR4*, *LGR5* and *LGR7* or a mimetic thereof, and
20 detecting the presence of a binding event between said receptor and ligand in said sample.

>LGR4 nucleotide sequence (SEQ ID NO:01)

ATGCCGGGCCCGCTAGGGCTGCTCTGCTTCCTCGCCCTGGGGCTGCTCGGCTCGGGCGGGCCACGCGCGCGGCCCGCCT
 CTCTGCGCGCGCCCTGCGACTGCGACGCGCGACCGTGGGTGGACTGCTCGGAAAGGGGTGACGCGCGCTACCGGAGGGT
 CTAGCGGCTCTACCCCAAGCACTGGATATCATAGTATGAACAATATCACCCAGTACGAGAAGATGCATTTAAAGATTTCOCA
 TTTCTAGAGGAGCTACAAGCTCGCTGGTACGACCTTTCTCTTATCATCTCAAAAGCGCTTGTCTGGCTGAAAGAACTCAAA
 GTCTCAACACTGCAAGAATACGTTGAGTGAACAGTGCCCACTGAAGSCATTACGAGCACTGAGTCTTTGCGAGCTTTTACGC
 TTAGATGCCAACCATATTACCTCAGTCCCGGAGGACAGTTTGAAGGGCTTGCCAGTTAGCCCATCTGTGGCTGGATGAC
 AACAGCTTGACGGAAGTGGCCGCTGGCTCCCTCAGCAACCTGCCAACCTGCGAGGGCTGACCTTTGGCTCTCAACAACATC
 TCAAGCTCCCTGACTTCTGCTTTACCAACCTTTCAAGCTTGGTCTTGCTGACTGCATGAACATAAAATTTAAAGCGCTC
 AGTCAACACTGTTTGTATGGACTAGATAACCTGGAAACCTTGGACTTGAATTACAATTAATTTGGATAGTTTCTCAGGCT
 ATTTAAAGCCCTTCCGAGCCTTAAAGAGCTGGGATTTACAGTAATCTATTCTGTGTTTCTGATGGAGCATTTGGTGGT
 AATCCACTGCTAAGAACTATTCAATTTGTATGATAATCCTCTGTCTTTTGTGGGGAATCAGACTTTCAACAACCTGTCTGAT
 CTGCTATGCTTAGTCACTTCTGTGTGCAAGCTGGTGAGTGGTTCGCCAATCTGACCGGAACCTGTCCATTGGAGAGTCTA
 ACCTTGACAGGACAAAAATAAGCAGCATACCTGATGATCTGTGCCAAAACAAAAGATGCTGAGGACTCTGGACTTATCT
 TATAACAATATAAGACCTTTCCAAGTTTAAATGGTGTCTGTCGATTGGAAGAAATTTATTGACGCGTAATCAAAATCTCC
 CTAATAAAGGAAATACTTTTCAAGGCTTAACATCTCTAAGGATTCTAGATCTGAGTAGAAACCTGATCCGTAATTTAC
 AGTGGAGCTTTTGCAGAGCTTGGGACAAATTAACCTGGATGTAAGTTTCAATGAATTAATTTCTTCCACGGAAGGC
 CTAATGGGCTCAATCAACTAAAGCTTTGGGTAACCTCAAGCTGAAGACGCTTGGCAGCCAGAGACTTTGCTAATCTC
 AGGTCTCTATCAGTACCATATGCTTATCAGTGTGTGCAATTTTGGGGGTGTGACTTTTATGCAAAATAAACACAGAAGAT
 AACAGCCCCAAGAACACAGTGTGACAAAGAGAAAGGTGCTACAGATGCAGCAAAATGTACACGACACTGCTGAGAAGCAA
 GAACATAGCCAAATAATTACCACTGTACACCTTCAACAGGTGCTTTCAAGCCCTGTGAATATTACTGGGAAGTCTGGATG
 ATTCGCTTACAGTGTGGTTCATTTCTCGTGGTCCGCTTGTCTTTCAACCTGCTGTGCTTTTAAACAGTGTGTGGCTGTGT
 TCATCACTGCTGCTCCAAACTCTTCATAGGCTTGATTTCTGTGCTAACTTACTCATGGGCATCTATCTGGCATCTCT
 ACTTTTCTGATGCTGTCTGCTGGGGCCGATTGGCCGAATTTGGCATTTTGGTGGGAACTGGCAGCGGCTCGAAGGTAGCC
 GGTCTCTGGAGCTTCTCTCTCAGAGAGCGCTGTATTTCTATTAACACTGCGAGTCTGGGAAGAAGCGTATTTGCAAG
 GATTGTGATGAAACAGGGGAAGAGCAGTCACTTCAGACAGTCCAGGTGGCCGCCCTCTTAGCTTTTCTGGGTGGCCGAGTG
 CGAGGCTGCTTCCCCCTTTTCCAGCGAGGGCAATATTCTGCATCGCCCTTGTGTTTGGCGTTTCTCAGAGGAGAAACCCCA
 TCGTTAGGATTCAGCTGTACCTTAGTGCTATTAACTCACTGGCAATTTTACTAATGGCCATTTATCTCACTAACTATAC
 TGCACCTTAGAGAAGGAGGAGCCTGTCGGAACCTCCAGCTTACGCGTGAATTAAGCAGTGTGCTGGCTCATCTTCAACAAC
 TGCATCTTCTTCTGCGCTTGTGCAATTTTCTCATTTGCAACATGATCAACGCAATCTCCATCAGCCCGGAGATAATGAAG
 TCTGTTACACTGATATTTCTCCCGTTGGCTGCTTGCCTGAATCCGCTCTGTATGTTTCTTCAACCCAAAGTTTAAAGAA
 GACTGGAAGCTACTGAAGCGCGCTGTTTACCAGGAAACAGGATCTGTTTCAAGTTTCCATCAGCAGCCAGCGGCTTGGGG
 GAACAGGATTTTCTACTATGACTGTGGCATGTATTTCCCACTTGCAGGGTAACCTGACTGTCTGTGACTGTCTGTGAGTCATT
 CTTTGTGACAAAACAGTATCATGCAAACTTAATAAAATCGCACAGTTGTCTGTATTGACAGCGGCTCTTGGCAGAGG
 CCGAGGGCTTACTGGTCTGATTGTGGTACAGACTCAGCCCATCTGACTACAGATGAAGAAGATTTCTTGTCTCAGAC
 AGCTCTGACCAAGTGCAGCGCTGTGGACGAGCCTGCTTCTACAGAGCTGCGGATTCCTCTGTGCGCTATGCTTACAAT
 CTACACAGAGTCAGAGACTGA

>LGR4 amino acid sequence (SEQ ID NO:02)

MPGFLGLLCLFALGLLGSAGPSGAAPPLCAAPCSCDGRDVRVDCSGKGLTAVPEGLSAFTQALDISMNNITQLPEDAFKSPF
 FLEELQAGNDLSLHPKALSGLEKLVLTQNNQLRTVPSEATHGLSALQSLRLDANHITSVPEDSFEGLVQLRHLWLDD
 NSLTVEPVRVNSLPTLQALTLALNNISSIPDFATNLSSVLVHLHNNIKLSQKCFADLDNLLETLDLNNYLVDEFPOA
 IKALPSLKELGFSNSISVIPDGAFFGNPLRLTHLYDNPLSFVGNISAFHNSLDLCLVIRGASLVGWQFNLITGVHLES
 TLGTGKISSIPDGLQKMLRLTLDLYNNIRLDPFNGCRALDEISLQRNQISLKLNETFOGLTSLRLDLNLRLEIRH
 SGAFKGLTITNLDVSNELTSFTEGLENLNLQKLVGNFKLDALARDFANLRSLSPVYAYQCCAFWGCDSCLKINTED
 NSPQESHVTEKGAIDANVTSIAENELSGSIQIIHCTPSTGAFKPCPEYLLGSWMIRLTWVFIPLVALLFNLLVITVFASC
 SSLPASKLFLIGLISVSNLLMGIYTGXLTFLDAVSWGRFAEFGIWWETGSGCKVAGSLVAFSSAEVLTLLAAVRSVFAK
 DLMKHGKSSHLRQFQVALLALLGAAGVCFPLFHGGQYASPLCLPFPPTGETPSLGTVTVLVLLNSLAFLLMAIYTKLY
 CLEKEDLSENSQSSVYKHAWLITFNCFIFCPVAFSFAPLITAISIPGKMSVLTIFPLPACLNPLVYVFFNPFKEK
 DWKLLKRRVTRKHGSYSVISISSQGGGEGDFYDCMGYSHLQGNLTWCDCESFLTKPVSCKHLIKSHSCPVLTAAACQR
 PEAYWSDCGTQSAHSYDAEEDSFVSDSSDQVACGRACFYQSGRFLVRYAYNLQRVRD

FIG. 1

>Nucleotide sequence of LGR5 (total 2082 nucleotides) (SEQ ID NO:03)

CTACATCTCCATMCAATAGAATCCACTCCCTGGGAAAGAAATGCTTTGATGGGCTCCACAGCCTAGAGACTTTAGATTTA
AATTACAAATACCTTGATGAATTTCCCACTGCAATTAGGACACTCTCCAACCTTAAGGAAGCTAGGATTTTCATAGCAACAAAT
ATCAGGTCGATAGCTGAGAAAGCATTTTGTAGGCAACCCCTCTCTTATTACAATACATTTCTATGACAAATCCCATCCCAATTT
GTTGGGAGATCTGCTTTTCAACATTTTACGTGAACCTAAGAACACTGACTCTGAATGGTGCTCTCAAAATACCTGAATTTCTCT
GATTTAACTGGAAGCTGCAAACTGGAGAGTCTGACTTTAACTGGAGCAGAGATCTCATCTCTCTCTCAAAACCGCTTGCAAT
CAGTTACTCAATCTCCAAGTGGCTAGATCTGCTCTTCAACCTATTAGAAGATTACCCAGTTTTTTCAGTCTGCGAAAGCTT
CAGAAATTTGACCTAAGACATATGAAATCTACGAAATTAAGTTGACACTTTCCAGCAGTTTGCTTAGCCTCCGATCGCTG
AATTTGGCTTTGGAACAAAATTTGCTATTATTCAACCCCAATGCAATTTTCCACTTTGGCACTCCCTTAATAAAGCTGGACCTATCG
TCCAACCTCCTGTCGCTCTTTTCTATAACTGGGTTACATGGTTTAACTCACTTAAATTAACAGGAAATCATGCTTACAG
AGCTGGATATCATCTGAAAACCTTCCGAACTCAAGGTATAGAAATGCCTTATGCTTACCAAGTCTGTGCTATTGGAGTG
TGTGAAGATGCCTATAAGATTCTTAATCAATGGAATTAAGGTGACCAACAGCAGTATGGACGACCTTCATAAGAAAGATGCT
GGAATGTTTCAAGCTCAAGATGAACGTGACCTTGAAGATTCTCTGCTTGACTTTGAGGAAGACCTGAAAGCCCTTCAATTC
GTGCACTGTTCACTTCCCGAGGCCCTCTCAAAACCTGTGAACACCTTGCTTATGAGTCTGCTGATCAGAATTTGAGGTGGT
ACCATAGCAGTTCTGGCACTTACTTGAATGCTTTGGTGACTTCAACAGTTTTCAGATCCCTCTGTGACATTTCCCCATT
AAACTGTTAATTTGGGTCATCGCAGCAGTGAACATGCTCAACGGAGTCTCCAGTCCGCTGCTGGCTGGTGGATGCGTTT
ACTTTTGGCAGCTTTGACGACATGCTGCTGGTGGGAGAAATGGGGTTGGTTGCCATGTCTTGGTTTGTCCATTTT
GCTTCAGAAATCATCTGTTTTCCTGCTTACTCTGGCAGCCCTGGAGCGTGGTTCTCTGTGAAATATTTCGCAAAATTTGAA
ACGAAAGCTCCATTTTCTAGCTGAAAGTAATCATTTTGTCTCTGTGCCCTGCTGGCCTTGACCATGCGCGCAGTTCCCTCT
CTGGGTGGCAGCAAGTATGGCGCTCCCTCTCTGCTGCTTTGCTTTTGGGAGCCGACGACCATGGGCTACATGCTC
GCTCTCATCTTGTCTCAATTTCCCTTTGCTTCTCATGATGACCATTTGCTACACCAAGCTCTACTGCAATTTGGACAAGGA
GACCTGGAGAATATTGGGACTGCTTATGGTAAACACATTTGCCCTGTGCTCTTCAACCACTGCATCTTAACTGCCCT
GTGGCTTTCTGTGCTTCTCTCTTAAATAAACCTTACATTTATCAGTCTCAAGTAATTAAGTTTATCTCTCTGCTGGTGA
TGCCCACTTCTGCTGCTCAATCCCTTCTCTACATCTTGTTCATCTCACTTTAAGGAGCTTGGTGAGCGCTGAGA
AAGCAAACTTACGCTGCGACAAGATCAAAACACCCCAAGCTTGATGTCAATTAACCTGATGATGTGAAAAACAGTCTGT
GACTCACTCAAGCTTGGTAACTTTTACAGCTCCAGCATCACTTATGACTGCTCCAGTCCCGTCCGATCACCAGCT
TATCCAGTGACTGAGAGTGCCATCTTCTCTGTGGCATTGTGCCATGTCTCTAA

>amino acid sequence of LGR5 (total 693 amino acids) (SEQ ID NO:04)

LHLHNNRIHSLGKKCFDGLHSLETLDLNNYNLDEFPTAIRTLNLKELGFHSNNIRISPEKAFVGNPSLITIHFDNPIQF
VGRSAFOHLPBLRLTLNGASQITEFPLDTGTANLESILTGAQISSLPQTVCNQLPNLQVLDSLNNLEDLPFSFVCOQL
QKIDLHNEIYEIKVDTPQQLSLRLNLAWNKIALIHPNAPSTLPSLIKLDLSNNLLSFPITGLHGLTHLKLGTNHALQ
SWISSENFPPELKVIMFPYAYQCCAFGVCENAYKISNOWNKGDNSMDDLHKDKDAGMFOAQDERDLEFLDPEEDLKALHS
VQCSPPGPFKPEHLLDGWLIRIGVWTIAVALTLCNALVTSTVFRSPLYISPIKLLGVIAAVNMLTGVSAAVLGVDFA
TFGSPARHGAWWENGVCVHIFGLSIFASESSVFLTLAALERGFVSKYAFKETFAPFSLLKVIILCALLATMAAVPL
LGGSKYGASPLCLPLPFGEPSSTMGVMVALILLNSLCFLMMTIAYTKLYCNLDKDGLENIDWDCSMVKHIALLLFTNCILNCP
VAFLSFSSLLNLTISPVEVILFLLVVVPLPACLNPLLYLLFNPHFKEDLVSLRKQTYVWTRSKHPSLMSINSDDVEKQSC
DSTQALVTFTHSSITYDPLPPSSVPSAPYTESCHLSVAVFVCL

FIG. 2

AAAAGGAGGGAAGAAAAAAGAGGAATGGAAAGAGACAGAGAAAGGAAATGGAGCTGGAAGGAGGGAGGACTGCTTT
 GTAACTGCTTAAGATTGCAGACAGAAATAGCACACCAACCTCTGAGCTGTATGTCAGATTTCAGAAAGCAAGACCAAAAT
 TTGCTCATCTTTCATTAAATCAGTTGTGTCAGATGAGAGGAAATGACATCTGGTCTGCTTCTTCATCATTTTAATTTT
 TGGAAAAATATTTTCTCATAGGGGGTGACAGAGTGTCTCAAGTGCTCTCTTGGTCAATTTCCCTCTGGGACATACAAA
 AGTCTCTGGCTCAGCTCTGCATCTGTCAGTGTAAAGCGTTGTGACAGCTCTGGGGAATCAGGCCGATGAGACACATGTGAGAGC
 AAGCAATGGATGGTTCATCCGAATTTGACAAATATTTTGGCAGTTACTACAAATAGTCTTCCCAATATCTTTTGGAG
 AGAACACCTGAAATGTTTGTGGCGTTGTGCGCACTGCAATGTCTTTTGGCAAGGTCTGGAGCTTGACTGTGTAGAAA
 CCAATTTACAGCGCTGTTCATCGGTTTCTTCAAATGTGACTGCGAATGCTACTTCAGTGGACATTAATAGAAAGACTG
 CTTCTCGATTTGCTTCAGAAATATCATGATCTCAGAAGCTCAATCTGCAAAACATAGAAATCATCAATCTCCAT
 CTATGCTCTCAGAGGACTGAATAGCCTTACTAAATGTTATCTCAGTCAATACAGCAATAACCTTCTGGAAGCGGST
 TTTTGAAGATCTCTCAGACATAGAAATGCTGTAATTAAGAAATATCCCTCAGTCGAATTTCCCAACCAACATTT
 TATGGACTAAATCTCTTATTCTCTTAGTCTGTGATGAATAACGTCCTCACCCGTTAGCTGTATAACCTCTCTGTCA
 ACACATGCCGAGACACTACATTGGTGGAGCTGTGAAGGCAACCATATCATAAATTAAGAAATTTGACTTTTATTCTCT
 CGAGTAAATTTAACTGTTTATGAGTAGGAGAAAAACAAATTAATCACTTAATAGAAAATCTTTTGCACTCTCCAG
 ATCTCGGATGAATATTTGGATTAGGAAGTAAATGAGATGAAATCTTCAACCGCTTAATTCAGGAATCAAGAGAGCT
 GAAACATAATGAATTTCTTCTATATCTCAATCCAGAAATTAACAAGTCCCAATTTGATTAATCTGTGCAACTCAAGT
 CTCTCAGCGCTGAGAAGGATGTAAATTTCAATATCCAAAGAGGTGTGATGACCTCTATGAATCTCTCATACATA
 TATTTTAAAGAAATTCAGATCTGCGGTTATGACACATGCTGTGCGAGCTGAACCAACCAAGTATGGAATTTATCT
 CTCAGAAAGATCTCTGGACAGCATTTATCAGAGAGTTATTTGTCTGGTGTGATTCGCAATCTTCTGTTGAAACA
 TTTTGTGATCTGCTATCGACCTTATATCAGAGTTCTGAGAACAGCTGGTATGCCATGTCAATCACTTCTCTCTGCTG
 CGCGACTCTTAAATGGGAATATATTTATCTGTGATCGGAGGCTTGACCTAAAGTTCTCTGGAGATACATAAAGCA
 TCGCGAGCTGTGATGGAGAGTACTCAATGTACGCTTTGTAGATCTTTGGCCATCTCTGCAACAGATATCAGTTT
 TACTGTTAACATTTCTGACGTTGGAAAAATACATCTGCATTTGCTGATCTCTTTTAGATGTGTGAGACCTGAAATATGC
 AGAACAAATTCAGTCTCTGATCTTCAATTTGGATATGCTGTTTATAGTGGCTTTCACTCAATTTAGACAAATAGGAATTT
 TTTTAAAAAATCATATGSCACCAATGGAGTAGTCTTCCCTCTCATTCAGAGATACAGAAATTTGGAGCCGAGA
 TTTATTTCAAGTGGCAATTTTCTTGTGATTAATTTGGCGCATTTATCATCATAGTTTCTTCTATGGAAAGCATGTTT
 TATAGTGTCTCATCAAGTGGCATACACAGCAACTGAAATACGGAATCAAGTTAAAAAAGAGATGATCTCTTGCCAAAG
 TTTTCTTTTATGATTAATCTGATGATCATATGCTGATACCAATTTTGTGTAGTGAAATTTCTTCAAGCTCTCTG
 TAGAATACAGGTAACATAACTCTCTGGGTAGTGTATTTATCTGCGCAATACAGTCTTTTGACGCCAATCTC
 TATACTCTGACACAGAGCAATTAAGAAATGATTCATCGGCTTTTGGTATACACTACAGCAAGAAATATCTGTGA
 CAGCAATGCTCAGAAAACATATGCTCATATCTCTGGTGAAATGTGCGCATCTGAGGATGCGCACTGTAGT
 TAATGAAGCGGACCTTTTACATACCTCTGTAAAGGTCTCACTCTCAATCAAGGACTCAATTCCTATTCA

MTSGSVFFYLILKPGKFSYSGGGQDVCKSLVPGFCPGNITKCLPQLHNNGVNDGCGQADENCGNPNNGMSGDFKFA
SYKMTSGQYFFEAE PCLGVGSVPVQCLQGLELDCTNLRNAPSVSNNLMSQWLNLIRKLPLDCKFKYMHDLKQ
LYLQNNKTSITSISYAFRGLNSLTKLYLSHNRIITFLPKGVNDLHRLWEILLIEDNHLNISIPPTVGYLNSILLVFWN
NVLTIRLDPKPLCOHMPRLHLWLDLEGNHINLITLTPGSCNLTVLVMKKNKHNHLENTPAQLKCLDELDLGNKIE
NLPLPILFKDKLKSQNLNSYNPIQKQANQFDYLVKLVLSLEGIESITNQKRMFLPMLNLSHYFKFKQCYGAAFP
SVSKCNTDGISSLENLNASIIQRVFWVVSATFCGNIFVICMMPYIRSEKNKLVMASISLCCADCLMGLIYLFVIG
GDFLKFSGEYKNHAQWLWMESTHCQLVGSLLALSIEVSVLLTFLTLKCYITVYFFRPGKRCITLILTLWTIG
FVIAPIPLSGEYFKFNKYGTGVCPLHSEDESIGASYSVAIFPLNLAAIPITVFSVSGMSFYVSHQASAITATE
RNOVKWEMILAKRFPIVFTDLACWPIPFVFKLVLSLOVEIPGTTISVNSYVFIPLDINSALNLYTLTTRPKEMIH
RFWYNRYQRKMSDMSKQGYAPSFIVEMVQLOEMPELMLKPDLTPTVPCMSLISQNSLRNYS*

>Final LGR7 (LGR7-Short variant) full length sequence (3584 nt)(SEQ ID NO:07)

CTGCTTTGTAACTGCTAAGATTGTCAGACAGAAATAGCACACAACCACTGTCGAGCTGTATGCGATTTCAGAAACCAAGA
 CCAAAATTTTGTCTCACTTTTCAATTAATCAGTTGCTCAGATAGAGGAAATGACACTGGTTTGTCTCTCTTCTATCACTCT
 TAATTTTGGAAAATATTTTCTCATGGGGGTGGACAGGATGTCAAGTGTCTCCCTTGGCTATTTCCCTGTGGGAAC
 ATCACAAGGTGCTTGCCTCAGCTCCTGCACGTGTAAACGGTGTGGAACGAGCTGCGGGAATCAGGCCGATGAGGACAACG
 TGTGGTGGTTTGTGCGCAGTGCATGTCTTTGGCAGGTCTGGAGCTTGACTGGATGAACCATTTACGAGATGTTCCAT
 CGGTTCTTCAAATGTGACTGCATGTCTCACTCAGTGGAACTTAATAAGAAAGCTTCTCTGATTTGCTTCAAGAAT
 TATCATGATCTTCAGAACTGGAGCTGCCAAAACAATAAGATTACATCCATCTCCATCTATGCTTTCAGAGGACTGAA
 TAGCCTTACTAACTGTATCTCAGTCATAACAGAATAAACCCTTCTCGAAGCCGGGTGTTTTGAAGATCTTTCACAGAC
 TAGAATGSGCTGATAATTGAAGATAAATCACTCAGTCGAAATTTCCCAACACATTTTATGAGACTAAATCTCTTATT
 CTCTTAGCTCTGATGAATACGCTTCTCACCCTGTTAACTGATAAACCTCTCTGTCAACACATGCCAAGACTACATTG
 GCTGGACCTTGAAGGCCAACCATATCCATAATTTAAGAAATTTGACTTTTATTTCTCCTGCAGTAATTTAACTCTGTTTATG
 TGAATGAGGAAAAACAAAATTAATCACTTAATGAAATACCTTTTGACCTCTCCAGAAACTGGATGAATTTGGATTTTATG
 GGAAGTAATAGATGAAAAATCTTCCACCGCTTATATTCAAGGACCTGAAGGAGCTGTCACAATTGAATCTTTCCTTA
 TAATCAATCCAGAAAATTCAGCAAAACCAATTTGATTATCTTGTCAAACCTCAAGTCTCTCAGCTAGAAAGGGATTG
 AAATTTCAAATATCCAAACAAAGGATGTTTAGACCTCTTATGAATCTCTCACAATATATTTAAGAAATTTCCAGTAC
 TGTGGGTATGCACCAACATGTTGCGAGCTGTAAACCAACACTGATGGAATTTTATCTCTAGAGAATCTTCTGGCAAG
 CATTATATCAGAGATATTTGTCTGGGTGTTATCTGCAGTTACCTGCTTTGGAACATCTTTTGTCAATTTGCTATCGCAG
 CTATATCAGGTCTGAGAACAGCTGTATGCCATGTCAATCACTTCTCTCTGTGCGCCAGCTGTTAATGGGAATA
 TATTTATTCGTGTGCGAGGCTTTGACTAAAGTTTCTGTGGAGATAACATAAGCATCGGCAGCTGTGGATGGAGAG
 TACTCACTTGCAGCTGTAGAGATCTTGGCCATTCTGTCACAGAAGTATCAGTTTACTCTTAACTATTTCTGACAT
 TGGAAAAATACATCTGCATTTCTCTATCTCTTTAGATGTGTGAGACCTGGAAAAATGCAGAACAAATACAGTTCTGATT
 CTATCTTGGATTTACTGGTTTTATGTAGTGGCTTTCATTCCATTGAGCAATAAGGAAATTTTCAAAAACACTATCTGGCAC
 CCACTGGAGATGCTCTCCCTCTCACTCAGAGAATACAGAAAGTATTTGGAGCCAGATTTTATCAGTGGCAATTTTCT
 TTGGTATTAATTTGGCCGCACTTATCATCATAGTTTTCCTATGGAAGCATGTTTATGATGTGTTCACTAAAGTGGCC
 ATAAACGACAACTGAAATACGGAATCAAGTTAAAAAAGAGATGATCCTTGCACAAACGTTTTCTTCTATAGATTATTC
 TGATGCTATTATGCTGTGATACCCATTTTTGTAGTGAATTTCTTCTACTGCTTCAGGTAGAAATACAGAGTACCAATA
 CCTCTTGGGTAGTGTATTTTATCTGCCCATTAACAGCTGCTTTGAACCAATCTCTATCTCTGACACACAGACA
 TTTTAAAGAAATGATCTCATCGTTTGTGTAATAACTACAGACAGAAAGAAATCTATTGAGCAGCAAGAGGTGAGAAACATA
 TGCTCCCTCATCTCATCTGGGTGGAATGTGGCCACTGCAGAGATGCGACCTGAGTTAATGAAGCCGAGCTCTTTTCA
 CATACCCCTGTGAAATGTCACTGATTTCTCAATCAACGAGACTCAATTTCTATTCATGACTGACTCTGAAATCTCAIT
 TCTTCGAGAGAATCACTGTGGGGGCTCTCATGAGGATTTTACTGGTATGAAATGAATACCAACAAAAATTAATTTAT
 AATTAATGCTAAGATAAATATTTTACAGGACATGAGGAAAAATAAAATGACATAATGCTCTTCAAGAGGAAGTAA
 TTATATCAATAAGTATATATATATTGATAGACATTTTGCATAAGAAATTAAGAGAAATCTACTTCAAGTAACATCTCAIT
 CATTTTTCTAACATGCACTTATTTAGTACCACATCTATGTGCATAGCATTGCAATATGATCTCTGGAAGTACAGCAT
 GCAGAACCTTTCAATCTGTAGATAGTGTAAATGACAAAGACATATCAAAAGTCCATCTGCAAGTTTCTGATTTAAAG
 TAGAGCTTTTACCTGTGCATCTGACCTCAGCAGAAATCATAGGCATTTTAAATAAAGGTTTAAAGTTTGGAAATCTCA
 GTGTATTTGCATCATAGAAAAATGTCTGACTGTTTGCACAAAATATATTTCTGTTTAAAGAAATCCATCTACTCTCTCTT
 AAGTTTTCATACACTCTGAGAGGCCAACACACATATTTTAACTAAAAAGAGTGTCTGTGATAGAACTCAAAAACAGCA
 CTCTCTTTTGGCAGCTTGTGCCAGTTTTCTCTTGTCTTAAATGAAACATCATATATGGAATTTGGAATAGGAGAGTA
 TGATGACGCGCAGAGAAGTGGATCAGAAAAACATAGAATGAGGATAAACATTTTCAATTTAGTGGAAACTCTGAAATTA
 TCCTTGATATGTCAGTTAACTGATTTTCAACAGGATGCCAAGCAAAAAGGCTTTTCAACAAACCGTCTGTTTTA
 AGAACAGCACTAAAGTGGTTTTAACTACCCCACTTATGAGGCTGAATGTATGTGTTGTCGAAATATCTCAGTAAAGCA
 GTTAAAGGAAAAAGAGCTGGAATGCACTGATGAGGAACCTTAATTCAGGAAGGAAAGGCTGTGATGTACACATTT
 CACTTTAAGCAGAAATCTTTCTTCAAGAAATGACTTTTACTTTCTCTGTCAGCTGCCAGCAGCTGAGATCTAACTTT
 TTTAATAGTTGTTCTTCTCTGAGTCTACGTTATATAGNATTTTTGCTTTCTCAATATGTGAACCTTTAAGCAGGAG
 AAAAAATTTTTCAGATGTTTCAAAATACNCAAAAAATTTTGCACACAAAAATCTGGAATCNAACCATAATG
 CCTTATTGAATATAGTTGTATAGNTTTGTTCTGAAAAACC

>Final LGR7-S ORF (722 amino acids) (SEQ ID NO:08)

MTSGSVFFYILIPGKYFHSHGQDVKSLGYFPCGNITKCLPQLLHCNGVDDCGNQADENCVVLVQCMSLPGLLEL
 DWMKPFVSVPSNSVNTAMSLQWNILRLPLDPCKFNYHDLQDLQNNKITSISIYAFRLNSLTKLYLSHNRITFL
 KPGVFEDLHRLWLIIEDNLNLSIPPTFYGLNSLILLVLMNNVLRLPDKPLCQHMPLRLHLLDGBNHIHNLNLT
 FISCSNLTVLVMRKNIHNLNLSIPAPLQKLELDDLSGNKINENLPLFKDLKELSQLNSYNP IQKIQKLPDFYLV
 KLKLSLLEGIEISNIQQRMPFRLMNLSHIYFKKFQYCYGAYPHVRKSKNPNTDGLSSLENLNASIIQRVFWVVSAVTC
 FKNIFVIMRPFYRSENKLYAMSIISLCCADCLMGIYLFVIGFDLKRFGYENKHAQWLMESTHCQLVGSAILSTE
 VSVLLLTFLTLKLEYICIVVPRCVRPGKRTITVLILWI TGFIVAFIPLSNKEKPFKNYGTGFCWCPHLLDEFTESI
 GAQIYISVAIPGLINLAAPIIIVPSYSGMFYSVHOSATATEIRINQVKEMILAKRFNIVFTDALCVIPFVKFLS
 LLOVEIPGTITSVVVIIFILPINSALNPILYLTTRPFKEMIHFRFWNYRQRKMSDKGQKYTAPSFIVWEMVPLQEM
 PPELMKPDFLFTYPEMSLISQSTRLSYS*

5/8

>Alignment of LGR7-L with LGR7-S

Query=LGR7-L

Sbjct=LGR7-S

```

Query: 1   MTSGSVFFYILIFGKYFSHGGGQDVKCSLGYFFPCGNITKCLPQLLHCNGVDDCGNQADED 60
           MTSGSVFFYILIFGKYFSHGGGQDVKCSLGYFFPCGNITKCLPQLLHCNGVDDCGNQADED 60
Sbjct: 1   MTSGSVFFYILIFGKYFSHGGGQDVKCSLGYFFPCGNITKCLPQLLHCNGVDDCGNQADED 60

Query: 61  NCGDNNGWSMQFDKYFASYKMTS QYPFEAETPECLVGSVPVQCCLCQ---GLELDCDET 117
           NC                               V V C C      GLELD  +
Sbjct: 61  NC-----VVVLQCQMSLPLGLELDWMKP- 82

Query: 118  LRAVPSVSSNVTAMSLQWNLIRKLPPDCFKNYHDLQKLQYLNQNKITSISIAFRGLNSLT 177
           +VPSVSSNVTAMSLQWNLIRKLPPDCFKNYHDLQKLQYLNQNKITSISIAFRGLNSLT 177
Sbjct: 83  FTSVPSVSSNVTAMSLQWNLIRKLPPDCFKNYHDLQKLQYLNQNKITSISIAFRGLNSLT 142

Query: 178  KLYLSHNRITFLKPGVFEDLHRLEWLIIEDNHLRSISPPTFYGLNSLILVLMMNVLTRL 237
           KLYLSHNRITFLKPGVFEDLHRLEWLIIEDNHLRSISPPTFYGLNSLILVLMMNVLTRL 237
Sbjct: 143 KLYLSHNRITFLKPGVFEDLHRLEWLIIEDNHLRSISPPTFYGLNSLILVLMMNVLTRL 202

Query: 238  PDKPLCQHMPRLHWLDLEGNHINHLRNLTIFISCSNLTVLVMRKNKINHNLNENTFAPLQKL 297
           PDKPLCQHMPRLHWLDLEGNHINHLRNLTIFISCSNLTVLVMRKNKINHNLNENTFAPLQKL 297
Sbjct: 203 PDKPLCQHMPRLHWLDLEGNHINHLRNLTIFISCSNLTVLVMRKNKINHNLNENTFAPLQKL 262

Query: 298  DELDLGSKNIENLPLIFKDLKELSQLNLSYNPIQIKIQAQFDYLVKLKSLSLLEGIEISN 357
           DELDLGSKNIENLPLIFKDLKELSQLNLSYNPIQIKIQAQFDYLVKLKSLSLLEGIEISN 357
Sbjct: 263 DELDLGSKNIENLPLIFKDLKELSQLNLSYNPIQIKIQAQFDYLVKLKSLSLLEGIEISN 322

Query: 358  IQQRMFRPLMNLSHIYFKKFQYCGYAPHVRSCKPNTDGISSLENLLASIIQRVFWVWVSA 417
           IQQRMFRPLMNLSHIYFKKFQYCGYAPHVRSCKPNTDGISSLENLLASIIQRVFWVWVSA 417
Sbjct: 323 IQQRMFRPLMNLSHIYFKKFQYCGYAPHVRSCKPNTDGISSLENLLASIIQRVFWVWVSA 382

Query: 418  VTCFGNIFVICMRPYIRSENKLYAMSIISLCCADCLMGIYLFVIGGFDLKFRGEYNKHAQ 477
           VTCFGNIFVICMRPYIRSENKLYAMSIISLCCADCLMGIYLFVIGGFDLKFRGEYNKHAQ 477
Sbjct: 383 VTCFGNIFVICMRPYIRSENKLYAMSIISLCCADCLMGIYLFVIGGFDLKFRGEYNKHAQ 442

Query: 478  LMMESTHCQLVGSAILSTEVSULLTFLTLEKYICIVYFRCVRPGKCRITITVLILWI 537
           LMMESTHCQLVGSAILSTEVSULLTFLTLEKYICIVYFRCVRPGKCRITITVLILWI 537
Sbjct: 443 LMMESTHCQLVGSAILSTEVSULLTFLTLEKYICIVYFRCVRPGKCRITITVLILWI 502

Query: 538  TGFIVAFIPLSNKEFFKNYYGTNGVCFPLHSEDTESIGAIYSVAIFLGINLAAFIIIVF 597
           TGFIVAFIPLSNKEFFKNYYGTNGVCFPLHSEDTESIGAIYSVAIFLGINLAAFIIIVF 597
Sbjct: 503 TGFIVAFIPLSNKEFFKNYYGTNGVCFPLHSEDTESIGAIYSVAIFLGINLAAFIIIVF 562

Query: 598  SYGSMFYSVHQSAITATEIRNQVKKEMILAKRFFFIVFTDALCWIPIFVVKFPLSLQVEI 657
           SYGSMFYSVHQSAITATEIRNQVKKEMILAKRFFFIVFTDALCWIPIFVVKFPLSLQVEI 657
Sbjct: 563 SYGSMFYSVHQSAITATEIRNQVKKEMILAKRFFFIVFTDALCWIPIFVVKFPLSLQVEI 622

Query: 658  PGTITSWVVFILPINSALNPILYTLTRPFKEMIHFRFWYNYRQRKSMDSKGQKTYAPSF 717
           PGTITSWVVFILPINSALNPILYTLTRPFKEMIHFRFWYNYRQRKSMDSKGQKTYAPSF 717
Sbjct: 623 PGTITSWVVFILPINSALNPILYTLTRPFKEMIHFRFWYNYRQRKSMDSKGQKTYAPSF 682

Query: 718  IWVEMWPLQEMPELMPKPDLYTTPCEMSLISQSTRLNSYS 757
           IWVEMWPLQEMPELMPKPDLYTTPCEMSLISQSTRLNSYS
Sbjct: 683 IWVEMWPLQEMPELMPKPDLYTTPCEMSLISQSTRLNSYS 722

```

FIG. 5

6/8
FIG. 6

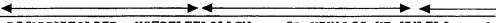
Signal peptide

LGR4 MPGPLGLLCFLALGLGSAGPSGA
LGR5 MDTSLGLVLLSLPVLLQLATG
LHR MKQRFSAQLQLKLLLLQPPLPRA
FSHR MALLLVSLLAFLSLGSG
TSHR MRPADLLQLVLLLDLFDLGG

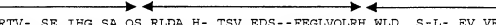
N-flank cysteine-rich sequence

LGR4 APPL AA-P S DGDR---RVD SGKGLTAVPEGLSAFTQA
LGR5 GSSPRSGVLLRG P-TH H EPDGRMLLRVD SDLGLSELPSNLSVFTSY
LHR LREAL P-EP N VPDG--ALR-- PGPTAGLTR
FSHR HHRI H SNRVFL---- QESKVTETPSDLPRNAIE
TSHR MG SSPP E HQEED--FRVT KDIQRIPLSPSTQT

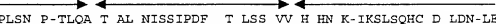
Leucine-rich repeats



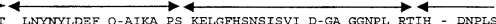
LGR4 DISMNNITQLPED KSPFFLEELQLAGN -- SL HPKALSG KE KVLTLQ -- Q
LGR5 DLSMNNISQLLNPPLPSLHFLLEELRLAGNA-- TY PKGA TG YS KVLMLQ -- Q
LHR SLAYLVPKVIQSQ RGLNEVIKIEISQI S- ER EANA DN LN SEILIQ TK -
FSHR RFVLTKLRVIQKG SGFGDLEKIEISQN V- EV EADV SN PK HEIRIEKAN -
TSHR KLIETHLRTIPSH SNLPNISRIYVSI- VT QQLESMS YN SKVTHIER TR -



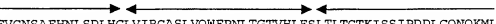
LGR4 RTV- SE IHG SA QS RLDA H- TSV EDG--FEGVLQRLH WLD S-L- EV VR
LGR5 RHV- TE LQN RS QS RLDA H- SYV P-SC-FSGLHSLRH WLD A-L- E VQ
LHR RYIE -G FIN PG KY SIC- TG RKF DVTKVFSSSESNTF- EIC LHI- T GN
FSHR LYIN -E FON PN QY LIS- TG KHL DVHK-IHSLQKVL- DIQ INIH - ERN
TSHR TYID -D LKE PL KF GIF- TGLKMF DLTK-VYSTDIFFI EIT PYM- S VN



LGR4 PLSN P-TLQA T AL NISSIPDF T LSS VV H HN K-IKSLSQHC D LDN-LE
LGR5 A RS S-ALQAMT AL KIHHPIDY G LSSWVV H HN R-IHSLGKKC D LHS-LE
LHR A QGMNNESTV K YG GFEEVQSH - GTT TS E KE VHLEKMHNGA R A-TGPK
FSHR S VG SFESVI W NK GIQEIHCN - GTQ DE N SD NNLEELPNDV H A-SGPV
TSHR A QG CNETLT K YN GFTSVQGY - GTK DAVY NK KYLTVIDKDA G VYSGPS



LGR4 T LNYNYLDEF Q-AIKA PS KELGFHSNSISVI D-GA GGNPL RTIH - DNPLS
LGR5 T LNYNNLDEF T-AIRT SN KELGFHSNNIRSI E-KA VGNPS ITIHF- DNPIQ
LHR T ISSTKLQAL SYGLESIQ R I-ATS-SYSLKKL SRET V-N- LEAT T ----
FSHR I ISRTRIHSL SYGLEN KK R-ARSTYN-LKKL TLEKLVA--- MEAS T ----
TSHR L VSQTSVTAL SKGLEH KE I-ARNTWT-LKKL LSLs LH--- TRAD S ----



LGR4 FVGNSAFHNLSDLHCLVIRGASLVQWFPNLTGTGVHLESLLTGTKISSIPDDLCONQXML
LGR5 FVGRSAFQHLPELRLTLNGASQITEFPDLTGTANLESLLTGAQISSLPQTVCNQLPNL
LHR -----
FSHR -----
TSHR -----

7/8

LGR4 RTLDLSYNNIRDLPSPNGCRALKEEISLQRNQISLIKENTFQGLTSLRILDLNRNLIREIH
 LGR5 QVLDLSYNLLLEDLPSPFVQCQLKIDLRHNEIYEIKVDTFQQLSLRSLNLAWNKAIAIH
 LHR -----
 FSHR -----
 TSHR -----

LGR4 SGAFAKLGTITNLDVSPNELTSFFTEGLNGLNQLK
 LGR5 PNAFSTLPGLIKLDLSSNLLSSFFPITGLHGLTHLK
 LHR -----
 FSHR -----
 TSHR -----

C-flank cysteine-rich sequence

LGR4 LVGNFKLKDALAARDFANLRSLV YAYQ WCDCSLCKLNTEDNSPQEHSVTKEGKA
 LGR5 LTGNHALQSLISSNFPELKVIEM YAYQ GVCENAYKISNQWNKGDNSMDDLHKK
 LHR ----- --SH RNLPTKEQNFHSHISENFSKQCESTVR
 FSHR ----- --SH ANWRRQISELHPICNKSILRQEVVDMT
 TSHR ----- --SH KNQKKIRGILESMLCNESSMQSLRQRK

LGR4 TDAANVTSTAENE HS-----
 LGR5 DAGMFQAQDERDL DF-----
 LHR KVSNTLYSSMLA SE-----
 FSHR QTRGQRSSLAEDN SS-----
 TSHR SVNALNSPLHQEY ENLGDSIVGYKEKSKFQDTHNNAHYVVFEEQEDEIIGFGQELKNP

LGR4 -----QIIIH T STGA K YLLGSWMI
 LGR5 -----LLDFEEDLKALHSVQ S SPGP K HLLDGNLI
 LHR -----LSGWDVEYQFCLPKTR- A EPDA N DIMGYDFL
 FSHR YSRGFDMTYTEFDYDLCEVVDVT S KPDA N DIMGYNIL
 TSHR QEETLQAFDSHYDYTCGDSSEDMV T KSDE N DIMGYKFL

Transmembrane

	TM 1										TM 2															
LGR4	LTV	F	FLV	LLF	LL	ILTVFA	CSS	PASKL	FIGLISV	SNLLM	ITYT	GILTFL	AVSW													
LGR5	IGV	T	AV	LTC	AL	TSTVFR	PLYIS	PIKL	IGVIA	AVNMLT	VSSAVL	G	AF F													
LHR	VLI	L	NI	IMG	MT	LFVLLT	RYK	TVPRF	MCNLS	FADFCM	LYLLLI	S	SQ K													
FSHR	VLI	F	SI	ITG	II	LVILIT	QYK	TVPRF	MCNLA	FADLCI	IYLLLI	S	IH K													
TSHR	IVV	FVSL	LLG	VF	LLILIT	HYK	NVPRF	MCNLA	FADFCM	MYLLLI	S	LY H														

	TM 3																
LGR4	GRFA	EFG	W	E	S	KV	SLA	S	SA	FL	LA	AV	S	VFAK	DLMKHKG	KSSH	QF
LGR5	GSFA	RHGW	AW	EN	V	HVI	LSI		S	FL	LAA			GFSV	KYSARFET	APFSSL	
LHR	QYYN	HA	D	Q	S	ST	FT		L	YT	VIT			WHTITY	AIHLQ	LR	HA
FSHR	SQYH	NYA	D	Q	A	DA	FT		L	YT	AIT			WHTITH	AMQLDC	VQ	HA
TSHR	SEYH	NYA	D	Q	P	NT	FT		L	YT	VIT			WYAIT	FAMRLDR	IR	HA

FIG. 6 (CONT)

8/8

	<u>TM 4</u>					<u>TM 5</u>			
LGR4	QVAALLALLGA	AVAGCF	PHGGQ	SASPL	FPTGETPSL	GFTVT	LV	SL	LLMA
LGR5	KVIILLCALL	ALTM	AV	L G K GASPL	LPFGEP	STM	G	MVALIL	SLC LMMT
LHR	ILIMLGGW	L	FSSLI	ML V V N MKVSI	F	MDVET	TL	SQV	IL
FSHR	ASVMVMGW	IFAF	AA	LF IF I S MKVSI		MDIDS	PL	SQ	VMSLLV
TSHR	CAIMVGGV	WCCFL	LL	V I S AKVSI	MDTET	PL	ALA	IV	FVLT

	<u>TM 6</u>			
LGR4	II T L	CNL-EKED	LENSQSSVI	HV W NCIFFC
LGR5	IA T L	CNL-DKGD	LENIW CSMV	HI L L NCILNC
LHR	AC I I	FAVRN	PELMATNK	TKIA KM I DFTCMA
FSHR	GC IHI	LTVRN	PNIVSSSS	TRIA RM M DFLCMA
TSHR	CCHV I	ITVRN	PQYNPGDK	TKIA RM V DFCMA

	<u>TM 7</u>			
LGR4	M SVTLI	F LPA	L V	VF N
LGR5	I FI LVVV	LPA	L L	IL N
LHR	S VL VL	Y INS	A F	AI T
FSHR	A IL VL	H INS	A F	AI T
TSHR	S IL VL	Y LNS	A F	AI T

C-terminal tail

LGR4	PK KE WKL KRRVTRKHG	SVSVSISSQGGCGEQDFYYDCGMYSHLQGNLTVCCCESFL
LGR5	PH KE LVS RKQTYVWTRSKHPSLMS	INSDVEKQSCDSTQALVFTTSSSITVDLPPSS
LHR	KT QR FFL LSKFGCCRRRAEL	YRRKDFSAYTNSCKNGFTGSKNPSQSTLKLSTLHCQG
FSHR	KN RR FFI LSKCGCYEMQAQI	YRTETSSTVHNTHPRNGHCSSAPRVTNGSTVILVPLS
TSHR	KA QR VFI LSKFGICKRQAQAYRGQ	RVPPKNSTDIQVQKVTHDMRQLHNMEDVYELI
LGR4	LTKPVSCKHILIKSHSCPV	LTAA SCQRPEAYWSDCGTQSAHSDYADEEDSFVSDSSDQVQA
LGR5	VPSPAYPVTE	SCHLSSVAFVPCL
LHR	TALLDKTRY	TEC
FSHR	HLAQN	
TSHR	ENSHLTPKKQGQ	ISEEYMQTVL
LGR4	CGRACFYQSRGFPLV	RYAYNLQVRD

FIG. 6 (CONT)

As a below-named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name,

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

Novel Mammalian G-Protein Coupled Receptors Having Extracellular Leucine Rich Repeat Regions
the specification of which (check only one item below):

☐ is attached hereto.

☐ was filed as United States application

Serial No. _____

on _____

and was amended

on _____ (if applicable).

☒ was filed as PCT international application

Number PCT/US99/06573

on 25 March 1999

and was amended under PCT Article 19

on _____ (if applicable).

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, §1.56(a).

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate or of any PCT international application(s) designating at least one country other than the United States of America listed below and have also identified below any foreign application(s) for patent or inventor's certificate or any PCT international application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) of which priority is claimed:

PRIOR FOREIGN/PCT APPLICATION(S) AND ANY PRIORITY CLAIMS UNDER 35 U.S.C. 119:

COUNTRY (If PCT, indicate "PCT")	APPLICATION NUMBER	DATE OF FILING (day, month, year)	PRIORITY CLAIMED UNDER 35 USC 119
US	60/079,501	26 March 1998 (26/03/98)	<input checked="" type="checkbox"/> YES <input type="checkbox"/> NO
			<input type="checkbox"/> YES <input type="checkbox"/> NO
			<input type="checkbox"/> YES <input type="checkbox"/> NO
			<input type="checkbox"/> YES <input type="checkbox"/> NO
			<input type="checkbox"/> YES <input type="checkbox"/> NO

Combined Declaration For Patent Application and Power of Attorney (Continued)

(Includes Reference to PCT International Applications)

ATTORNEY'S DOCKET NUMBER

STAN-084

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) or PCT international application(s) designating the United States of America that is/are listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in that/those prior application(s) in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, §1.56(a) which occurred between the filing date of the prior application(s) and the national or PCT international filing date of this application:

PRIOR U.S. APPLICATIONS OR PCT INTERNATIONAL APPLICATIONS DESIGNATING THE U.S. FOR BENEFIT UNDER 35 U.S.C. 120:

U.S. APPLICATIONS		STATUS (Check one)		
U.S. APPLICATION NUMBER	U.S. FILING DATE	PATENTED	PENDING	ABANDONED

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith. (List name and registration number)

Karl Bozicevic, Reg. No. 28,907

Bret E. Field, Reg. No. 37,620

Dianna L. DeVore, Reg. No. 42,484

Alan W. Cannon, Reg. No. 34,977

Carol L. Francis, Reg. No. 35,513

Pamela J. Sherwood, Reg. No. 36,877

Paula A. Borden, Reg. No. 42,344

Nicola Verona, Reg. No. P47,153

Send Correspondence to:

BOZICEVIC, FIELD & FRANCIS LLP

200 Middlefield Road, Suite 200

Menlo Park, CA 94025

Telephone: (650) 327-3400

Facsimile: (650) 327-3231

Direct Telephone Calls to:
(name and telephone number)Name: Bret E. Field
Registration No. 37,620
Telephone: (650) 327-3400

201	FULL NAME OF INVENTOR	FAMILY NAME HSUEH	FIRST GIVEN NAME AARON	SECOND GIVEN NAME J.W.
	RESIDENCE & CITIZENSHIP	CITY STANFORD CA	STATE OR FOREIGN COUNTRY CALIFORNIA	COUNTRY OF CITIZENSHIP UNITED STATES
	POST OFFICE ADDRESS	POST OFFICE ADDRESS 25 RYAN COURT	CITY STANFORD	STATE & ZIP CODE/COUNTRY CALIFORNIA 94305/US
202	FULL NAME OF INVENTOR	FAMILY NAME HSU	FIRST GIVEN NAME SHEAU	SECOND GIVEN NAME YU
	RESIDENCE & CITIZENSHIP	CITY MOUNTAIN VIEW CA	STATE OR FOREIGN COUNTRY CALIFORNIA	COUNTRY OF CITIZENSHIP TAIWAN
	POST OFFICE ADDRESS	POST OFFICE ADDRESS 234 ESCUELA AVENUE, APT. 40	CITY MOUNTAIN VIEW	STATE & ZIP CODE/COUNTRY CALIFORNIA 94040/US
203	FULL NAME OF INVENTOR	FAMILY NAME LIANG	FIRST GIVEN NAME SHAN-GUANG	SECOND GIVEN NAME
	RESIDENCE & CITIZENSHIP	CITY PALO ALTO CA	STATE OR FOREIGN COUNTRY CALIFORNIA	COUNTRY OF CITIZENSHIP CHINA
	POST OFFICE ADDRESS	POST OFFICE ADDRESS 438 VENTURA AVENUE #9	CITY PALO ALTO	STATE & ZIP CODE/COUNTRY CALIFORNIA 94306/US

FULL NAME OF INVENTOR	FAMILY NAME <u>VAN DER SPEK</u>	FIRST GIVEN NAME <u>PETRUS</u>	SECOND GIVEN NAME <u>JOHANNES</u>
RESIDENCE & CITIZENSHIP	CITY <u>QSS NLX</u>	STATE OR FOREIGN COUNTRY <u>NETHERLANDS</u>	COUNTRY OF CITIZENSHIP <u>NETHERLANDS</u>
POST OFFICE ADDRESS	POST OFFICE ADDRESS <u>BREMLAAN 10</u>	CITY <u>OSS</u>	STATE & ZIP CODE/COUNTRY <u>5342 HM OSS/NL</u>

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

SIGNATURE OF INVENTOR 201 <i>[Signature]</i>	SIGNATURE OF INVENTOR 202 <i>[Signature]</i>	SIGNATURE OF INVENTOR 203 <i>[Signature]</i>
DATE <u>SEP. 11, 2000</u>	DATE <u>9-11-00</u>	DATE <u>9-17-00</u>
SIGNATURE OF INVENTOR 204		
DATE		

M:FORMS\COMDEC.WP 5/96

005250 20040902

COMBINED DECLARATION FOR PATENT APPLICATION AND POWER OF ATTORNEY
(Includes Reference to PCT International Applications)

ATTORNEY'S DOCKET NUMBER

STAN-084

As a below-named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name,

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

Novel Mammalian G-Protein Coupled Receptors Having Extracellular Leucine Rich Repeat Regions
the specification of which (check only one item below):

☐ is attached hereto.

☐ was filed as United States application

Serial No. _____

on _____

and was amended

on _____ (if applicable).

☒ was filed as PCT international application

Number PCT/US99/06573

on 25 March 1999

and was amended under PCT Article 19

on _____ (if applicable).

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, §1.56(a).

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate or of any PCT international application(s) designating at least one country other than the United States of America listed below and have also identified below any foreign application(s) for patent or inventor's certificate or any PCT international application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) of which priority is claimed:

PRIOR FOREIGN/PCT APPLICATION(S) AND ANY PRIORITY CLAIMS UNDER 35 U.S.C. 119:

COUNTRY (if PCT, indicate "PCT")	APPLICATION NUMBER	DATE OF FILING (day, month, year)	PRIORITY CLAIMED UNDER 35 USC 119
US	60/079,501	26 March 1998 (26/03/98)	<input checked="" type="checkbox"/> YES <input type="checkbox"/> NO
			<input type="checkbox"/> YES <input type="checkbox"/> NO
			<input type="checkbox"/> YES <input type="checkbox"/> NO
			<input type="checkbox"/> YES <input type="checkbox"/> NO
			<input type="checkbox"/> YES <input type="checkbox"/> NO

Combined Declaration For Patent Application and Power of Attorney (Continued)

(Includes Reference to PCT International Applications)

ATTORNEY'S DOCKET NUMBER

STAN-084

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) or PCT international application(s) designating the United States of America that is/are listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in that/those prior application(s) in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, §1.56(a) which occurred between the filing date of the prior application(s) and the national or PCT international filing date of this application:

PRIOR U.S. APPLICATIONS OR PCT INTERNATIONAL APPLICATIONS DESIGNATING THE U.S. FOR BENEFIT UNDER 35 U.S.C. 120:

U.S. APPLICATIONS		STATUS (Check one)		
U.S. APPLICATION NUMBER	U.S. FILING DATE	PATENTED	PENDING	ABANDONED

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith. (List name and registration number)

Karl Bozicevic, Reg. No. 28,807
 Bret E. Field, Reg. No. 37,620
 Dianna L. DeVoro, Reg. No. 42,484
 Alan W. Cannon, Reg. No. 34,977

Carol L. Francis, Reg. No. 36,513
 Pamela J. Sherwood, Reg. No. 36,577
 Paula A. Borden, Reg. No. 42,344
 Nicole Verona, Reg. No. P47,153

Send Correspondence to:

BOZICEVIC, FIELD & FRANCIS LLP

200 Middlefield Road, Suite 200

Menlo Park, CA 94025

Telephone: (650) 327-3400

Facsimile: (650) 327-3231

 Direct Telephone Calls to:
 (name and telephone number)

 Name: Bret E. Field
 Registration No. 37,620
 Telephone: (650) 327-3400

201	FULL NAME OF INVENTOR	FAMILY NAME HSUEH	FIRST GIVEN NAME AARON	SECOND GIVEN NAME J.W.
	RESIDENCE & CITIZENSHIP	CITY STANFORD	STATE OR FOREIGN COUNTRY CALIFORNIA	COUNTRY OF CITIZENSHIP UNITED STATES
	POST OFFICE ADDRESS	POST OFFICE ADDRESS 25 RYAN COURT	CITY STANFORD	STATE & ZIP CODE/COUNTRY CALIFORNIA 94305/US
202	FULL NAME OF INVENTOR	FAMILY NAME HSU	FIRST GIVEN NAME SHEAU	SECOND GIVEN NAME YU
	RESIDENCE & CITIZENSHIP	CITY MOUNTAIN VIEW	STATE OR FOREIGN COUNTRY CALIFORNIA	COUNTRY OF CITIZENSHIP TAIWAN
	POST OFFICE ADDRESS	POST OFFICE ADDRESS 234 ESCUELA AVENUE, APT. 40	CITY MOUNTAIN VIEW	STATE & ZIP CODE/COUNTRY CALIFORNIA 94040/US
203	FULL NAME OF INVENTOR	FAMILY NAME LIANG	FIRST GIVEN NAME SHAN-GUANG	SECOND GIVEN NAME
	RESIDENCE & CITIZENSHIP	CITY PALO ALTO	STATE OR FOREIGN COUNTRY CALIFORNIA	COUNTRY OF CITIZENSHIP CHINA
	POST OFFICE ADDRESS	POST OFFICE ADDRESS 438 VENTURA AVENUE #9	CITY PALO ALTO	STATE & ZIP CODE/COUNTRY CALIFORNIA 94306/US

204	FULL NAME OF INVENTOR	FAMILY NAME	FIRST GIVEN NAME	SECOND GIVEN NAME
	RESIDENCE & CITIZENSHIP	CITY	STATE OR FOREIGN COUNTRY	COUNTRY OF CITIZENSHIP
	POST OFFICE ADDRESS	POST OFFICE ADDRESS	CITY	STATE & ZIP CODE/COUNTRY
		VAN DER SPEK	PETRUS	JOHANNES
		OSS	NETHERLANDS	NETHERLANDS
		BREMLAAN 10	OSS	5342 HM OSS/NL

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

SIGNATURE OF INVENTOR 201	SIGNATURE OF INVENTOR 202	SIGNATURE OF INVENTOR 203
DATE 18 Sept 2000	DATE	DATE
SIGNATURE OF INVENTOR 204		
DATE		

M:\FORMS\COMDEC\12 5/96

005250 29024950

SEQUENCE LISTING

<110> Hsueh, Aaron
Hsu, Yu Sheau
Liang, Shan-Guang
van der Spek, Petrus Johannes

<120> Novel Mammalian G-Protein Coupled
Receptors Having Extracellular Leucine Rich Repeat Regions

<130> SUN-84PCT

<160> 8

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 2856
<212> DNA
<213> human

<400> 1

atgcggggcc	cgctagggct	gctctgcttc	ctcgcccttg	ggctgctcgg	ctcgccgggg	60
cccagcgggc	cggcgcggcc	tctctgcgcg	gcgccttgca	gctgcgacgg	cgaccgtcgg	120
gtgactgctc	ccggaaggga	gttgacggcc	gtaccggagg	gtctcagcgc	cttcacccaa	180
gcactggata	tcagtatgaa	caatatcacc	cagttaccag	agatgcatt	taagagtttc	240
ccattttcag	aggagctaca	actggctggt	aacgaccttt	ctcttatcca	tccaaaggcc	300
ttgtctgggc	tgaagaacct	caaagtccca	acactccaga	ataatcagtt	gagaacagtg	360
cccagtgga	ccatccacgg	actgagtgct	ttgcagtcct	tacgcttaga	tgcccaacct	420
attacctcag	tcccggaggga	cagttttgaa	gggcttgctc	agttacgcga	tctgtggctg	480
gatgacaaca	gcttgacgga	agtgcacctg	cgtccctcca	gcaacctgcc	aacctctcag	540
gcgctgacct	tggctctcaa	caacatccca	agcatccctg	actctgcttt	caccaacctt	600
tcaagctctg	tggttctgca	tctgcataac	aataaaatta	aaagcctcag	tcaacaactgt	660
tttgatggac	tagataaacct	ggaacacctg	gacttgaatt	acaattactt	ggatgagttt	720
ctccaggcta	ttaaagccct	tcccagcctt	aaagagctgg	gatttcacag	taattctatt	780
tctgttattc	ctgatggagc	atttggtggt	aatccactgc	taagaactat	tcatttgtat	840
gataatccct	tgtcttttgt	ggggaaactca	gcatttcaca	acctgtctga	tctgatttgc	900
ttagtcatcc	tgggtgcga	ctgggtgcag	tggttcccca	tctgacccg	aactgtccat	960
tggagagctg	taacctgtac	agggaacaaa	ataagcagca	atactgatga	tctgtgccaa	1020
aaccaaaaga	tgctgaggac	tctggacctta	tctataaaca	atataagaga	ccttccaaagt	1080
tttaatggtt	tgctgtcgtt	ggaagaaatt	tcattgcagc	gtaatacaat	ctccctaata	1140
aaggaaaata	cttttcaagg	cttaacatct	ctaaggattc	tagatctgag	tagaacaacctg	1200
atccgttgaa	tccacagctg	agcttttgca	aagcttggga	caattactaa	ctggatgta	1260
agtttcaatg	aattaaactc	atttccactg	gaaggcctaa	atgggctcaa	tcaactaaag	1320
cttgttggtg	acttcaagct	gaagacgccc	ttggcagcca	gagactttgc	taactctcag	1380
tctctatcag	taccatatgc	ttatcagttg	tgtgcatttt	gggggtgtga	ctctttatgc	1440
aaattaaaca	cagaagataca	cagcccccaa	gaacacagtg	tgacaaaaga	gaaggtgctg	1500
acagatgcag	caaatgtcac	cagcatcgtc	gagacagga	aacatagcca	aataattatc	1560
cactgtacac	cttcaacagg	tgttttcaag	cctgtgtaat	atttactggc	agactggatg	1620
attgcgctta	cagtgtgggt	caatttccgt	gtcgcttgc	tttttaacct	gctgtcatt	1680
ttacagtggt	ttgcgttcgt	tteatcactg	cctgcctcca	aactcttcac	agcttgatt	1740
tctgtgtcta	acttactcat	gggcactcat	actggcatcc	ttacttttct	tgatgctgtg	1800
tccgtggggc	gattttgcga	atttggcatt	tgggtggaaa	ctggcagcgg	ctgcaaggta	1860
gccgttggtt	tggcagttct	ctccctcagag	agcgtctgat	tccatttaac	actggcagct	1920
gtggaaagaa	gcgtattttc	aaaggatttg	atgaaacacg	ggaagagcag	taccctcaaga	1980
cagttccacg	tggcgcctcg	cttagctttg	ctgggtgcgc	cagttggcag	ctgcttcccc	2040
cttttccacg	gagggcaata	tctgcactcg	cccttgtgtt	tgccgtttcc	tacaggagaa	2100

```

accccatcgt taggattcac tgtgacctta gtgctattaa actcactggc atttttacta 2160
atggccatta tctacactaa actatactgc aacttagaga aggaggacct gtcggaaaac 2220
tcccagttcta gcgtgattaa gcacgttgcc tggctcatct tcacaaacct catcttcttc 2280
tgccctgttg catttttctc atttgaccca ttgatcacgg caatctccat cagcccccag 2340
ataatgaagt ctgttacact gatattcttc cegtggcctg cttgcctgaa tccggctcgt 2400
tatgttttct tcaacccaaa gttaaagaa gactggaagc tactgaagcg gcgtgtttacc 2460
aggaaacacg gatctgtttc agtttccatc agcagccaag gcggtttgtg ggaacacgat 2520
ttctactatg actgtggcat gtattccac ttgcagggtt acctgactgt ctgtgactgc 2580
tgtgagtcac ttcttttgac aaaaccagta tcatgcgaac accttaataaa atcgacagt 2640
tgctctgtat tgacagggcg ctcttgccag aggcagagag cctactggtc tgattgtgtg 2700
acacagtcag cccattctga ctatgcagat gaagaagatt cctttgtctc agacagctct 2760
gaccaggtgc aggcctgttg acgagctcgc ttctaccaga gtctgtggatt cctctgtgtg 2820
cgctatgctt acaatctaca gagagtcaga gactga 2856

```

<210> 2

<211> 951

<212> PRT

<213> human

<400> 2

```

Met Pro Gly Pro Leu Gly Leu Leu Cys Phe Leu Ala Leu Gly Leu Leu
1 5 10 15
Gly Ser Ala Gly Pro Ser Gly Ala Ala Pro Pro Leu Cys Ala Ala Pro
20 25 30
Cys Ser Cys Asp Gly Asp Arg Arg Val Asp Cys Ser Gly Lys Gly Leu
35 40 45
Thr Ala Val Pro Glu Gly Leu Ser Ala Phe Thr Gln Ala Leu Asp Ile
50 55 60
Ser Met Asn Asn Ile Thr Gln Leu Pro Glu Asp Ala Phe Lys Ser Phe
65 70 75 80
Pro Phe Leu Glu Glu Leu Gln Leu Ala Gly Asn Asp Leu Ser Ile Leu
85 90 95
His Pro Lys Ala Leu Ser Gly Leu Lys Glu Leu Lys Val Leu Thr Leu
100 105 110
Gln Asn Asn Gln Leu Arg Thr Val Pro Ser Glu Ala Ile His Gly Leu
115 120 125
Ser Ala Leu Gln Ser Leu Arg Leu Asp Ala Asn His Ile Thr Ser Val
130 135 140
Pro Glu Asp Ser Phe Glu Gly Leu Val Gln Leu Arg His Leu Trp Leu
145 150 155 160
Asp Asp Asn Ser Leu Thr Glu Val Pro Val Arg Pro Leu Ser Asn Leu
165 170 175
Pro Thr Leu Gln Ala Leu Thr Leu Ala Leu Asn Asn Ile Ser Ser Ile
180 185 190
Pro Asp Phe Ala Phe Thr Asn Leu Ser Ser Leu Val Val Leu His Leu
195 200 205
His Asn Asn Lys Ile Lys Ser Leu Ser Gln His Cys Phe Asp Gly Leu
210 215 220
Asp Asn Leu Glu Thr Leu Asp Leu Asn Tyr Asn Tyr Leu Asp Glu Phe
225 230 235 240
Pro Gln Ala Ile Lys Ala Leu Pro Ser Leu Lys Glu Leu Gly Phe His
245 250 255
Ser Asn Ser Ile Ser Val Ile Pro Asp Gly Ala Phe Gly Gly Asn Pro
260 265 270
Leu Leu Arg Thr Ile His Leu Tyr Asp Asn Pro Leu Ser Phe Val Gly
275 280 285
Asn Ser Ala Phe His Asn Leu Ser Asp Leu His Cys Leu Val Ile Arg
290 295 300

```

Gly Ala Ser Leu Val Gln Trp Phe Pro Asn Leu Thr Gly Thr Val His
 305 310 315 320
 Leu Glu Ser Leu Thr Leu Thr Gly Thr Lys Ile Ser Ser Ile Pro Asp
 325 330 335
 Asp Leu Cys Gln Asn Gln Lys Met Leu Arg Thr Leu Asp Leu Ser Tyr
 340 345 350
 Asn Asn Ile Arg Asp Leu Pro Ser Phe Asn Gly Cys Arg Ala Leu Glu
 355 360 365
 Glu Ile Ser Leu Gln Arg Asn Gln Ile Ser Leu Ile Lys Glu Asn Thr
 370 375 380
 Phe Gln Gly Leu Thr Ser Leu Arg Ile Leu Asp Leu Ser Arg Asn Leu
 385 390 395 400
 Ile Arg Glu Ile His Ser Gly Ala Phe Ala Lys Leu Gly Thr Ile Thr
 405 410 415
 Asn Leu Asp Val Ser Phe Asn Glu Leu Thr Ser Phe Pro Thr Glu Gly
 420 425 430
 Leu Asn Gly Leu Asn Gln Leu Lys Leu Val Gly Asn Phe Lys Leu Lys
 435 440 445
 Asp Ala Leu Ala Ala Arg Asp Phe Ala Asn Leu Arg Ser Leu Ser Val
 450 455 460
 Pro Tyr Ala Tyr Gln Cys Cys Ala Phe Trp Gly Cys Asp Ser Leu Cys
 465 470 475 480
 Lys Leu Asn Thr Glu Asp Asn Ser Pro Gln Glu His Ser Val Thr Lys
 485 490 495
 Glu Lys Gly Ala Thr Asp Ala Ala Asn Val Thr Ser Thr Ala Glu Asn
 500 505 510
 Glu Glu His Ser Gln Ile Ile Ile His Cys Thr Pro Ser Thr Gly Ala
 515 520 525
 Phe Lys Pro Cys Glu Tyr Leu Leu Gly Ser Trp Met Ile Arg Leu Thr
 530 535 540
 Val Trp Phe Ile Phe Leu Val Ala Leu Leu Phe Asn Leu Leu Val Ile
 545 550 555 560
 Leu Thr Val Phe Ala Ser Cys Ser Ser Leu Pro Ala Ser Lys Leu Phe
 565 570 575
 Ile Gly Leu Ile Ser Val Ser Asn Leu Leu Met Gly Ile Tyr Thr Gly
 580 585 590
 Ile Leu Thr Phe Leu Asp Ala Val Ser Trp Gly Arg Phe Ala Glu Phe
 595 600 605
 Gly Ile Trp Trp Glu Thr Gly Ser Gly Cys Lys Val Ala Gly Ser Leu
 610 615 620
 Ala Val Phe Ser Ser Glu Ser Ala Val Phe Leu Leu Thr Leu Ala Ala
 625 630 635 640
 Val Glu Arg Ser Val Phe Ala Lys Asp Leu Met Lys His Gly Lys Ser
 645 650 655
 Ser His Leu Arg Gln Phe Gln Val Ala Leu Leu Ala Leu Leu Gly
 660 665 670
 Ala Ala Val Ala Gly Cys Phe Pro Leu Phe His Gly Gly Gln Tyr Ser
 675 680 685
 Ala Ser Pro Leu Cys Leu Pro Phe Pro Thr Gly Glu Thr Pro Ser Leu
 690 695 700
 Gly Phe Thr Val Thr Leu Val Leu Leu Asn Ser Leu Ala Phe Leu Leu
 705 710 715 720
 Met Ala Ile Ile Tyr Thr Lys Leu Tyr Cys Asn Leu Glu Lys Glu Asp
 725 730 735
 Leu Ser Glu Asn Ser Gln Ser Ser Val Ile Lys His Val Ala Trp Leu
 740 745 750
 Ile Phe Thr Asn Cys Ile Phe Phe Cys Pro Val Ala Phe Phe Ser Phe
 755 760 765

Ala Pro Leu Ile Thr Ala Ile Ser Ile Ser Pro Glu Ile Met Lys Ser
 770 775 780
 Val Trp Leu Ile Phe Phe Pro Leu Pro Ala Cys Leu Asn Pro Val Leu
 785 790 795 800
 Tyr Val Phe Phe Asn Pro Lys Phe Lys Glu Asp Trp Lys Leu Leu Lys
 805 810 815
 Arg Arg Val Thr Arg Lys His Gly Ser Val Ser Val Ser Ile Ser Ser
 820 825 830
 Gln Gly Gly Cys Gly Glu Gln Asp Phe Tyr Tyr Asp Cys Gly Met Tyr
 835 840 845
 Ser His Leu Gln Gly Asn Leu Thr Val Cys Asp Cys Cys Glu Ser Phe
 850 855 860
 Leu Leu Thr Lys Pro Val Ser Cys Lys His Leu Ile Lys Ser His Ser
 865 870 875 880
 Cys Pro Val Leu Thr Ala Ala Ser Cys Gln Arg Pro Glu Ala Tyr Trp
 885 890 895
 Ser Asp Cys Gly Thr Gln Ser Ala His Ser Asp Tyr Ala Asp Glu Glu
 900 905 910
 Asp Ser Phe Val Ser Asp Ser Ser Asp Gln Val Gln Ala Cys Gly Arg
 915 920 925
 Ala Cys Phe Tyr Gln Ser Arg Gly Phe Pro Leu Val Arg Tyr Ala Tyr
 930 935 940
 Asn Leu Gln Arg Val Arg Asp
 945 950

<210> 3

<211> 2082

<212> DNA

<213> human

<400> 3

ctacatctcc	ataacaatag	aatccactcc	ctgggaaaga	aatgctttga	tgggctccac	60
agcctagaga	cttttagattt	aaattacaat	aacctgtatg	aattcccccac	tgcaattagg	120
acactctcca	acttaaaagga	actaggattt	catagcaaca	atatcaggtc	gatacctgag	180
aaagcatttg	taggcaaccc	ttctcttatt	acaatacatt	tctatgacaa	tcccatccaa	240
tttgtttgga	gatctgtctt	tcaacattta	cttgaactaa	gaacactgac	tctgaatggt	300
gcctcacaaa	taactgaatt	tcttgattta	actggaaactg	caaacctgga	gagtctgact	360
ttaactggag	cacagatctc	atctcttctc	caaaccgtct	gcaactcagt	acctaactctc	420
caagctgctag	atctgtctta	caacctatta	gaagatttac	ccagtttttc	agctcgccaa	480
aagcttcaga	aaattgacct	aagacataat	gaaatctacg	aaattaaagt	tgacactttc	540
cagcagttgc	ttagcctccg	atcgctgaat	ttgctttgga	acaaaaattgc	tattattacc	600
cccaatgcct	tttccacttt	gccatcccta	ataaagctgg	acctatcgtc	caacctccctg	660
tcgtcttttc	ctataactcg	gttacatggt	ttaactcact	taaaattaac	aggaaatcat	720
gccttacaga	gctggatatt	atctgaaaac	tttcagaaac	tcaaggtinat	agaaatgcct	780
tatgcttacc	agtgcgtgac	atttgaggag	tgtagaagag	ccataaagat	tcttaaatcaa	840
tggaataaag	gtgacaacag	cagtatggac	gaccttcata	agaaagatgc	tggaatgttt	900
caggctccaag	atgaaactga	ccttgaagat	ttcctgcttg	actttgagga	agacctgaaa	960
gcctctcatt	cagtgcaagt	ttcaccttcc	ccaggccctc	tcaaacctctg	tgaacacctg	1020
cttgatggct	ggctgatcac	aattggagtg	tggaacctag	cagttctggc	acttaactgt	1080
aatgcttttg	tgacttcaac	agttttcaaga	tccccctgtg	acattttccc	cattaaactg	1140
ttaattgggg	tcatcgagc	agtgaacatg	ctcacgggag	tctccagtg	cggtgctggct	1200
ggtgtggagt	cggtcacctt	tggcagcttt	gcacgacatg	tgctctgggt	ggagaatggg	1260
gttggttgcc	atgtcatgtg	ttttttgtcc	atttttgctt	cagaatcatc	tgttttcctg	1320
cttactctgt	cagccctgga	gcgtgggttc	tcttggaagt	attctgcaaa	atttgaacag	1380
aaagctccat	tttctagcct	gaaagtaact	attttgctct	gtgccctgct	ggccttgacc	1440
atggcccgag	ttccccctgt	gggtggcgag	aagtatggcg	ctctccccct	ctcgctggct	1500
ttgctctttg	gggaagccag	caaccatggc	tacatggtcg	ctctcatctt	gctcaattcc	1560
cttttgcttc	tcattgtgac	cattgcctac	accaagctct	actgcaattt	ggacaaggga	1620
gacctggaga	acatttgagg	ctgctctatg	gtaaaacaca	ttgccctgtt	gctcttcacc	1680

```

aactgcaccc taaactgccc tgtggcttcc ttgtccctct cctctttaat aaaccttaca 1740
tttatcagtc ctgaagtaat taagtttata ctcttggtgg tagtcccaact tccctgatgt 1800
ctcaatcccc ttctctacat ctgtttcaat cctcacttta aggaggatct ggtgagcctg 1860
agaaagcaaa cctacgtctg gacaagatca aaacacccaa gcttgatgtc aattaactct 1920
gatgatgtcg aaaaacagtc ctgtgactca actcaagcct tggtaacctt taccagctcc 1980
agcatcactt atgacctgcc tcccagttcc ggcacacac cagcttatcc agtgactgag 2040
agctgcacac ttctctctgt ggcatttgtc ccattgtctt aa 2082

```

<210> 4
 <211> 693
 <212> PRT
 <213> human

<400> 4

Leu	His	Leu	His	Asn	Asn	Arg	Ile	His	Ser	Leu	Gly	Lys	Lys	Cys	Phe
1				5					10					15	
Asp	Gly	Leu	His	Ser	Leu	Glu	Thr	Leu	Asp	Leu	Asn	Tyr	Asn	Asn	Leu
			20					25					30		
Asp	Glu	Phe	Pro	Thr	Ala	Ile	Arg	Thr	Leu	Ser	Asn	Leu	Lys	Glu	Leu
		35					40					45			
Gly	Phe	His	Ser	Asn	Asn	Ile	Arg	Ser	Ile	Pro	Glu	Lys	Ala	Phe	Val
		50				55					60				
Gly	Asn	Pro	Ser	Leu	Ile	Thr	Ile	His	Phe	Tyr	Asp	Asn	Pro	Ile	Gln
		65			70					75				80	
Phe	Val	Gly	Arg	Ser	Ala	Phe	Gln	His	Leu	Pro	Glu	Leu	Arg	Thr	Leu
			85					90					95		
Thr	Leu	Asn	Gly	Ala	Ser	Gln	Ile	Thr	Glu	Phe	Pro	Asp	Leu	Thr	Gly
			100				105						110		
Thr	Ala	Asn	Leu	Glu	Ser	Leu	Thr	Leu	Thr	Gly	Ala	Gln	Ile	Ser	Ser
			115				120					125			
Leu	Pro	Gln	Thr	Val	Cys	Asn	Gln	Leu	Pro	Asn	Leu	Gln	Val	Leu	Asp
			130			135						140			
Leu	Ser	Tyr	Asn	Leu	Leu	Glu	Asp	Leu	Pro	Ser	Phe	Ser	Val	Cys	Gln
			145			150				155				160	
Lys	Leu	Gln	Lys	Ile	Asp	Leu	Arg	His	Asn	Gln	Ile	Tyr	Glu	Ile	Lys
			165						170					175	
Val	Asp	Thr	Phe	Gln	Gln	Leu	Leu	Ser	Leu	Arg	Ser	Leu	Asn	Leu	Ala
			180					185					190		
Trp	Asn	Lys	Ile	Ala	Ile	Ile	His	Pro	Asn	Ala	Phe	Ser	Thr	Leu	Pro
			195				200					205			
Ser	Leu	Ile	Lys	Leu	Asp	Leu	Ser	Ser	Asn	Leu	Leu	Ser	Ser	Phe	Pro
			210			215					220				
Ile	Thr	Gly	Leu	His	Gly	Leu	Thr	His	Leu	Lys	Leu	Thr	Gly	Asn	His
			225			230				235				240	
Ala	Leu	Gln	Ser	Trp	Ile	Ser	Ser	Glu	Asn	Phe	Pro	Glu	Leu	Lys	Val
			245					250						255	
Ile	Glu	Met	Pro	Tyr	Ala	Tyr	Gln	Cys	Cys	Ala	Phe	Gly	Val	Cys	Glu
			260			265						270			
Asn	Ala	Tyr	Lys	Ile	Ser	Asn	Gln	Trp	Asn	Lys	Gly	Asp	Asn	Ser	Ser
			275				280					285			
Met	Asp	Asp	Leu	His	Lys	Lys	Asp	Ala	Gly	Met	Phe	Gln	Ala	Gln	Asp
			290			295					300				
Glu	Arg	Asp	Leu	Glu	Asp	Phe	Leu	Leu	Asp	Phe	Glu	Glu	Asp	Leu	Lys
			305			310					315			320	
Ala	Leu	His	Ser	Val	Gln	Cys	Ser	Pro	Ser	Pro	Gly	Pro	Phe	Lys	Pro
			325					330						335	
Cys	Glu	His	Leu	Leu	Asp	Gly	Trp	Leu	Ile	Arg	Ile	Gly	Val	Trp	Thr
			340				345						350		

Ile Ala Val Leu Ala Leu Thr Cys Asn Ala Leu Val Thr Ser Thr Val
 355 360 365
 Phe Arg Ser Pro Leu Tyr Ile Ser Pro Ile Lys Leu Leu Ile Gly Val
 370 375 380
 Ile Ala Ala Val Asn Met Leu Thr Gly Val Ser Ser Ala Val Leu Ala
 385 390 395 400
 Gly Val Asp Ala Phe Thr Phe Gly Ser Phe Ala Arg His Gly Ala Trp
 405 410 415
 Trp Glu Asn Gly Val Gly Cys His Val Ile Gly Phe Leu Ser Ile Phe
 420 425 430
 Ala Ser Glu Ser Ser Val Phe Leu Leu Thr Leu Ala Ala Leu Glu Arg
 435 440 445
 Gly Phe Ser Val Lys Tyr Ser Ala Lys Phe Glu Thr Lys Ala Pro Phe
 450 455 460
 Ser Ser Leu Lys Val Ile Ile Leu Leu Cys Ala Leu Leu Ala Leu Thr
 465 470 475 480
 Met Ala Ala Val Pro Leu Leu Gly Gly Ser Lys Tyr Gly Ala Ser Pro
 485 490 495
 Leu Cys Leu Pro Leu Pro Phe Gly Glu Pro Ser Thr Met Gly Tyr Met
 500 505 510
 Val Ala Leu Ile Leu Leu Asn Ser Leu Cys Phe Leu Met Met Thr Ile
 515 520 525
 Ala Tyr Thr Lys Leu Tyr Cys Asn Leu Asp Lys Gly Asp Leu Glu Asn
 530 535 540
 Ile Trp Asp Cys Ser Met Val Lys His Ile Ala Leu Leu Leu Phe Thr
 545 550 555 560
 Asn Cys Ile Leu Asn Cys Pro Val Ala Phe Leu Ser Phe Ser Ser Leu
 565 570 575
 Ile Asn Leu Thr Phe Ile Ser Pro Glu Val Ile Lys Phe Ile Leu Leu
 580 585 590
 Val Val Val Pro Leu Pro Ala Cys Leu Asn Pro Leu Leu Tyr Ile Leu
 595 600 605
 Phe Asn Pro His Phe Lys Glu Asp Leu Val Ser Leu Arg Lys Gln Thr
 610 615 620
 Tyr Val Trp Thr Arg Ser Lys His Pro Ser Leu Met Ser Ile Asn Ser
 625 630 635 640
 Asp Asp Val Glu Lys Gln Ser Cys Asp Ser Thr Gln Ala Leu Val Thr
 645 650 655
 Phe Thr Ser Ser Ser Ile Thr Tyr Asp Leu Pro Pro Ser Ser Val Pro
 660 665 670
 Ser Pro Ala Tyr Pro Val Thr Glu Ser Cys His Leu Ser Ser Val Ala
 675 680 685
 Phe Val Pro Cys Leu
 690

<210> 5

<211> 2467

<212> DNA

<213> human

<400> 5

gaaaggagga	aagaaaaaaa	gaggaatgga	aagagacaga	gaaaggaaaat	gggagtgga	60
ggaggaggga	ctgctttgta	actgctaaga	ttgcagacag	aaatagcaca	caaccactgt	120
gagctgtatg	cgattcagaa	accaagacca	aattttgctc	actttcatta	atcagttgct	180
cagatagaag	gaaatgacat	ctgggtctgt	cttcctctac	atcttaattt	ttggaanaat	240
ttttctcat	gggggtggac	aggatgtcaa	gtgctccctt	ggctatttcc	ccgtgggaa	300
catcacaag	tgcttgccct	agctcctgca	ctgtaacggt	gtggacgact	gcgggaatca	360
ggccgatgag	gacaaactgt	gagacaacaa	tggaatgtcc	atgcaatttg	acaaatattt	420
tgccagttac	tacaaaaatg	cttcccaata	tccttttgag	gcagaaacac	ctgaatgttt	480

ggctcggttct gtgcaccatgc aatgtctttg ccaaggtctg gagcttgact gtgatgaac 540
 caattttacga gctgttccat cggtttcttc aaatgtgact gcaatgtcac ttcagtgga 600
 cttaataaga aagcttctct ctgattgtct caagaattat catgatcttc agaagctgt 660
 cctgcacaaac aataagatta catccatctc catctatgct ttcagaggac tgaatagcct 720
 tactaaactg tatctcagtc ataacagaat aaccttctcg aagccgggtg tttttgaaga 780
 ttccacaga ctagaatggc tgataattga agataatcac ctcagtcgaa tttcccacc 840
 aacatthtat ggactaaatt ctctatttct cttagtcctg atgaataacg tctccaccgc 900
 tttaccctgat aaacctctct gtcaacacat gccaaagact cattggctgg accttgaag 960
 caaccatctc cataatttaa gaaatttgac ttttatttcc acttttgac ctctccagaa 1020
 agtgatgagg aaaaaacaaa ttaactacct aaatgaaaaa ctttttgac ctctccagaa 1080
 actggatgaa ttggatttag gaagtaataa gattgaaaaa acttttgac ctctccagaa 1140
 ggacctgaag gagctgtcac aattgaattct ttcctataat ccaatccaga aaattcaagc 1200
 aaaccaattt gattatcttg tcaaacctcaa gtctctcagc ctagaaggga ttgaaatttc 1260
 aaatatccaa caaaggatgt ttgacacctc tatgaattct tctcacatat attttgaaga 1320
 attccagtcg attcattctg caccacatgt tcgcagctgt aaaccaaaca ctgatggaat 1380
 ttcattctca gagaattctc tggcaagcat tatccagaga gtatttgtct ggggtgtatc 1440
 tgcagttacc gtctttgtaa acatttttgt catttgcatg cgacctata tcaggtctga 1500
 gaacaagctg tatgcatctc caatcatttc tctctgctgt gccgactgct taatgggaat 1560
 atattttatc gtgactggag gctttgacct aaagtttcgt ggagaataca atagcatgc 1620
 gcagctgtgg atggagagta ctcattgtca gcttgtagga tctttggcca tctgtgccac 1680
 agaagtatca gttttactgt taacatttct gacattggaa aaatacatct gcattgtcta 1740
 tctttttaga tgtgtgagac ctggaataat gacatacaat acagttctga tctcatttg 1800
 gattactggt tttatagtgg ctttcaatcc attgagcaat aaggaatttt tcacaaacta 1860
 ctatggcacc aatggagtat gcttccctct tcattcagaa gatacagaaa ctatgggagc 1920
 ccagatttat cgaatggcaa tttttcttgg tattaattgt gccgcatta tcatcatagt 1980
 tttttcttat tgaagcatgt tttatagtgt tcatcaaat gccataacag tcagggtaga 2040
 cgggaatcaa gttaaaaaag agatgatcct tgccaaaact tttttctta tagtatttac 2100
 tgatgcatta tgcctgatac ccatttttgt agtgaatttt ctgccatgc ttcaggtagt 2160
 gaataccaggt accataacct cttgggtagt gatttttatt ctgccatta acagtgcttt 2220
 gaaccaattt cctatactc tgaccacaag accattttaa gaaatgattc atcggttttg 2280
 gtataactac agacaaaaga aatctatgga cagcaaaact cagaaaacat atgtccctac 2340
 attcatctcg gtggaaatgt ggccactgca ggagatgcca ctgagtttaa tgaagccgca 2400
 ccttttcaaa taacctctgt aatgtcact gattttctca tcaacgagac tcaattctca 2460
 tcatga

<210> 6

<211> 757

<212> PRT

<213> human

<400> 6

Met	Thr	Ser	Gly	Ser	Val	Phe	Phe	Tyr	Ile	Leu	Ile	Phe	Gly	Lys	Tyr
1			5					10					15		
Phe	Ser	His	Gly	Gly	Gln	Asp	Val	Lys	Cys	Ser	Leu	Gly	Tyr	Phe	
			20				25					30			
Pro	Cys	Gly	Asn	Ile	Thr	Lys	Cys	Leu	Pro	Gln	Leu	Leu	His	Cys	Asn
			35				40					45			
Gly	Val	Asp	Asp	Cys	Gly	Asn	Gln	Ala	Asp	Glu	Asp	Asn	Cys	Gly	Asp
			50				55				60				
Asn	Asn	Gly	Trp	Ser	Met	Gln	Phe	Asp	Lys	Tyr	Phe	Ala	Ser	Tyr	Tyr
						70				75				80	
Lys	Met	Thr	Ser	Gln	Tyr	Pro	Phe	Glu	Ala	Glu	Thr	Pro	Glu	Cys	Leu
						85				90				95	
Val	Gly	Ser	Val	Pro	Val	Gln	Cys	Leu	Cys	Gln	Gly	Leu	Glu	Leu	Asp
						100				105				110	
Cys	Asp	Glu	Thr	Asn	Leu	Arg	Ala	Val	Pro	Ser	Val	Ser	Ser	Asn	Val
						115				120				125	
Thr	Ala	Met	Ser	Leu	Gln	Trp	Asn	Leu	Ile	Arg	Lys	Leu	Pro	Pro	Asp
						130				135				140	

Cys Phe Lys Asn Tyr His Asp Leu Gln Lys Leu Tyr Leu Gln Asn Asn
 145 150 155 160
 Lys Ile Thr Ser Ile Ser Ile Tyr Ala Phe Arg Gly Leu Asn Ser Leu
 165 170 175
 Thr Lys Leu Tyr Leu Ser His Asn Arg Ile Thr Phe Leu Lys Pro Gly
 180 185 190
 Val Phe Glu Asp Leu His Arg Leu Glu Trp Leu Ile Ile Glu Asp Asn
 195 200 205
 His Leu Ser Arg Ile Ser Pro Pro Thr Phe Tyr Gly Leu Asn Ser Leu
 210 215 220
 Ile Leu Leu Val Leu Met Asn Asn Val Leu Thr Arg Leu Pro Asp Lys
 225 230 235 240
 Pro Leu Cys Gln His Met Pro Arg Leu His Trp Leu Asp Leu Glu Gly
 245 250 255
 Asn His Ile His Asn Leu Arg Asn Leu Thr Phe Ile Ser Cys Ser Asn
 260 265 270
 Leu Thr Val Leu Val Met Arg Lys Asn Lys Ile Asn His Leu Asn Glu
 275 280 285
 Asn Thr Phe Ala Pro Leu Gln Lys Leu Asp Glu Leu Asp Leu Gly Ser
 290 295 300
 Asn Lys Ile Glu Asn Leu Pro Pro Leu Ile Phe Lys Asp Leu Lys Glu
 305 310 315 320
 Leu Ser Gln Leu Asn Leu Ser Tyr Asn Pro Ile Gln Lys Ile Gln Ala
 325 330 335
 Asn Gln Phe Asp Tyr Leu Val Lys Leu Lys Ser Leu Ser Leu Glu Gly
 340 345 350
 Ile Glu Ile Ser Asn Ile Gln Gln Arg Met Phe Arg Pro Leu Met Asn
 355 360 365
 Leu Ser His Ile Tyr Phe Lys Lys Phe Gln Tyr Cys Gly Tyr Ala Pro
 370 375 380
 His Val Arg Ser Cys Lys Pro Asn Thr Asp Gly Ile Ser Ser Leu Glu
 385 390 395 400
 Asn Leu Leu Ala Ser Ile Ile Gln Arg Val Phe Val Trp Val Val Ser
 405 410 415
 Ala Val Thr Cys Phe Gly Asn Ile Phe Val Ile Cys Met Arg Pro Tyr
 420 425 430
 Ile Arg Ser Glu Asn Lys Leu Tyr Ala Met Ser Ile Ile Ser Leu Cys
 435 440 445
 Cys Ala Asp Cys Leu Met Gly Ile Tyr Leu Phe Val Ile Gly Gly Phe
 450 455 460
 Asp Leu Lys Phe Arg Gly Glu Tyr Asn Lys His Ala Gln Leu Trp Met
 465 470 475 480
 Glu Ser Thr His Cys Gln Leu Val Gly Ser Leu Ala Ile Leu Ser Thr
 485 490 495
 Glu Val Ser Val Leu Leu Leu Thr Phe Leu Thr Leu Glu Lys Tyr Ile
 500 505 510
 Cys Ile Val Tyr Pro Phe Arg Cys Val Arg Pro Gly Lys Cys Arg Thr
 515 520 525
 Ile Thr Val Leu Ile Leu Ile Trp Ile Thr Gly Phe Ile Val Ala Phe
 530 535 540
 Ile Pro Leu Ser Asn Lys Glu Phe Phe Lys Asn Tyr Tyr Gly Thr Asn
 545 550 555 560
 Gly Val Cys Phe Pro Leu His Ser Glu Asp Thr Glu Ser Ile Gly Ala
 565 570 575
 Gln Ile Tyr Ser Val Ala Ile Phe Leu Gly Ile Asn Leu Ala Ala Phe
 580 585 590
 Ile Ile Ile Val Phe Ser Tyr Gly Ser Met Phe Tyr Ser Val His Gln
 595 600 605

Ser Ala Ile Thr Ala Thr Glu Ile Arg Asn Gln Val Lys Lys Glu Met
 610 615 620
 Ile Leu Ala Lys Arg Phe Phe Phe Ile Val Phe Thr Asp Ala Leu Cys
 625 630 635 640
 Trp Ile Pro Ile Phe Val Val Lys Phe Leu Ser Leu Leu Gln Val Glu
 645 650 655
 Ile Pro Gly Thr Ile Thr Ser Trp Val Val Ile Phe Ile Leu Pro Ile
 660 665 670
 Asn Ser Ala Leu Asn Pro Ile Leu Tyr Thr Leu Thr Thr Arg Pro Phe
 675 680 685
 Lys Glu Met Ile His Arg Phe Trp Tyr Asn Tyr Arg Gln Arg Lys Ser
 690 695 700
 Met Asp Ser Lys Gly Gln Lys Thr Tyr Ala Pro Ser Phe Ile Trp Val
 705 710 715 720
 Glu Met Trp Pro Leu Gln Glu Met Pro Pro Glu Leu Met Lys Pro Asp
 725 730 735
 Leu Phe Thr Tyr Pro Cys Glu Met Ser Leu Ile Ser Gln Ser Thr Arg
 740 745 750
 Leu Asn Ser Tyr Ser
 755

<210> 7
 <211> 3584
 <212> DNA
 <213> human

<400> 7
 ctgctttgta actgctaaga ttgcagacag aaatagcaca caaccactgt gagctgtatg 60
 cgattcagaa accaagacca aattttgctc actttcatta atcagttgct cagatagaag 120
 gaatgacat ctggttctgt ttctttctac attctaaatt ttgaaaaata tttttctcat 180
 ggggggtggac aggatgtcaa gtgctccctt ggcattttcc cctgtgggaa catcacaaa 240
 tgcttgcttc agctcctgca ctgtaacggt gtggacgact gcgggaatca gccgatgag 300
 gcaaacctgt tgggtgtttt gtgccagttc atgtctttgc caggctctgga gcttgactgg 360
 atgaaacctat ttacagagtg tccatcggtt tcttcaaatg tgaactgcaat gtcaactcag 420
 tggaaacttaa taagaaaagt tctcctcgat gtgttcaaga attatcatga tcttcagaag 480
 ctggacctgc aaaaacaataa gattacatcc atctccatct atgttttcag aggaactgaat 540
 agccttacta aactgtatct cagtcataac agaataacct tcttgaagcc ggggtgtttt 600
 gaagatcttc acagactaga atggctgata attgaagata atcacctcag tcgaatttcc 660
 ccaccaacat tttatggact aaattctctt attctcttag tcttgatgaa taacctcttc 720
 acccgttac ctgataaacc tctctgtcaa cacatgccaa gactacattg gctggacctt 780
 gaagccaacc atatccataa ttttaagaaat ttgactttta tttcctgcag taatttaact 840
 gttttagtga tgaggaaaaa caaaattaat cacttaaatg aaaaactctt tgcacctctc 900
 cagaaactgg atgaattgga tttaggaagt aataagattg aaaaactctc accgcttata 960
 ttcaaggacc tgaaggagct gtcaaatgt aatctttctt ataatacaat ccagaaaaat 1020
 caagcaaaccc aattttgatta tctttgcaaa ctcaagtctc tcagctaga agggattgaa 1080
 atttcaataa tccaacaaag gatgtttaga cctcttatga atctctctca catatatttt 1140
 aagaatttcc agtactgtgg gtatgcacca agcattatc agagagtatt tgtctggggt 1200
 ggaatttcat cctagagaaa tctcttgcca atgttgcga gctgtaaaac aaacactgat 1260
 gtatctgcag ttacctgtct tggaaaacatt tttgtcattt gcatgcgaga ctgcttaag 1320
 tctgagaaca agctgtatgc catgtcaact attctctctt tctgtgcaga atacaataag 1380
 ggaatatatt tttctgtgat cggagcgttt gacctaaagt ttctgtgaga atcaataaag 1440
 catgcccagc tgtggatgga gactactcat tgtcactgt taggatcttt ggcactctgt 1500
 tccacagaag tatcagtttt actgttaaca tttctgacat tggaaaaata catctgcatt 1560
 gctctatcct tttagatgtg gagacctgga aatgacagaa caattacagt tctgatcttc 1620
 atttgattga ctgggtttat agtggctttc attccattga gcaataaagg atttttcaaa 1680
 aactactatg gcaccaatgg agtatgcttc cctcttcact cagaagatgc agaaagtatt 1740
 ggagcccgaa tttattcagtg ggcaattttt cttggtatta atttggcgcc atttatcatc 1800
 atagtttttt cctatgggaag catgttttat agtgttcatc aaagtgcact acacagcaact 1860
 gaaataacgga atcaagttaa aaaaagagatg atccttgcca aacgtttttt ctttatagta 1920

```

tttactgatg cattatgctg gataccaccatt ttgtagtga aatttctttc actgcttcag 1980
gtagaataac caggtaccat aacctcttgg gtagtgattt ttattctgcc cattaacagt 2040
gctttgaacc caattctcta tactctgacc acagaccatt ttaagaaat gatcatcgg 2100
ttttggtata actacagaca aagaaaatct atggacagca aaggtcagaa aacatatgct 2160
ccatcaattca tctgggtgga aatgtggcca ctgcagagga tgcacactga gtaatatgag 2220
ccggaccctt tcacataccc ctgtgaaatg tcactgattt ctcaatcaac gagactcaat 2280
tcctattcat gactgacctt gaaattcatt ttcttcgaga gaatactgtg ggggtgcttc 2340
atgagggttt tactgggatg aaaatgaata ccccaaaatt aatttataat aatagctaag 2400
ataaaatttt tacaaggaca tgaggaaaaa taataatgac taatgctctt acaaagggaa 2460
gtaattatat caataatgta tatatatagg tagacatttt gcataagaaa ttaagagaaa 2520
tctacttcag taacattcat tcatttttct aacatgcatt tatgagtagc coactactat 2580
gtgcatagca ttgcaatata gtcttggaag tagacagtgc agaactcttc aactctgtag 2640
tagtgtttaa tgacaaaaga ctatacaaa gcatcctgca ttctctgtag taaagttaga 2700
ctttactctg catgtgcacg agcaagaatc ataggcaact ttaataaaag gtttaaaagt 2760
ttggaataact cagtgttatt gcatcataga aaatgtctga ctgtttgcga aataatattc 2820
tggttttaaga atccatctta cctctcttta agtttccata cacttgagag ccaacacaac 2880
atatttatta ctaaaaagat gctttgctag aaactcaaaa acagcacttc ttttggaact 2940
tcctgcccag ttttctcttt gctttaaatg aacatcatca tatggaattg gaatagaga 3000
gtatgagtag ggcagagaa ggtatcagaa aaactagaat gaggataaac atttacatta 3060
gtggaacttc ctgaaataaa tccttgattt gtcagttaac tgattttcaa caaggatgcc 3120
aagacaaaaa ggcttttcaa caaacctgtc tgttttaaga acagacctaa gtggtttaat 3180
tcaccacttt tagatgggtg aatgttatgg tgtgtgaaat atctcagtaa agcaggttaa 3240
agtaaaaaa gctggaatgc actgattcag gaacttaatt tcaggaaagg aaggtctgta 3300
tgatcacact tcactttaag cagaaaaatc ttcttcaaga aatgactctta ctttctcttt 3360
gcactgcag cactgagatg actaactttt taactagtgt ttcttctcta gtctctacgt 3420
tattagnatt ttttctttc ataattgtaa acctttaagc aggagaagaa aatgttttca 3480
gatagtttca aatacnccaa aaatgtttgc aacacaaaaa tactggaatc naaccataat 3540
gcccttaatt aatatatagt tgtatagnnt tgttctgaaa accc 3584

```

<210> 8

<211> 722

<212> PRT

<213> human

<400> 8

```

Met Thr Ser Gly Ser Val Phe Phe Tyr Ile Leu Ile Phe Gly Lys Tyr
1 5 10 15
Phe Ser His Gly Gly Gln Asp Val Lys Cys Ser Leu Gly Tyr Phe
20 25 30
Pro Cys Gly Asn Ile Thr Lys Cys Leu Pro Gln Leu Leu His Cys Asn
35 40 45
Gly Val Asp Asp Cys Gly Asn Gln Ala Asp Glu Asp Asn Cys Val Val
50 55 60
Val Leu Cys Gln Cys Met Ser Leu Pro Gly Leu Glu Leu Asp Trp Met
65 70 75 80
Lys Pro Phe Thr Ser Val Pro Ser Val Ser Ser Asn Val Thr Ala Met
85 90 95
Ser Leu Gln Trp Asn Leu Ile Arg Lys Leu Pro Pro Asp Cys Phe Lys
100 105 110
Asn Tyr His Asp Leu Gln Lys Leu Asp Leu Gln Asn Asn Lys Ile Thr
115 120 125
Ser Ile Ser Ile Tyr Ala Phe Arg Gly Leu Asn Ser Leu Thr Lys Leu
130 135 140
Tyr Leu Ser His Asn Arg Ile Thr Phe Leu Lys Pro Gly Val Phe Glu
145 150 155 160
Asp Leu His Arg Leu Glu Trp Leu Ile Ile Glu Asp Asn His Leu Ser
165 170 175
Arg Ile Ser Pro Pro Thr Phe Tyr Gly Lys Leu Asn Ser Leu Ile Leu Leu
180 185 190

```

Val Leu Met Asn Asn Val Leu Thr Arg Leu Pro Asp Lys Pro Leu Cys
 195 200 205
 Gln His Met Pro Arg Leu His Trp Leu Asp Leu Glu Gly Asn His Ile
 210 215 220
 His Asn Leu Arg Asn Leu Thr Phe Ile Ser Cys Ser Asn Leu Thr Val
 225 230 235 240
 Leu Val Met Arg Lys Asn Lys Ile Asn His Leu Asn Glu Asn Thr Phe
 245 250 255
 Ala Pro Leu Gln Lys Leu Asp Glu Leu Asp Leu Gly Ser Asn Lys Ile
 260 265 270
 Glu Asn Leu Pro Pro Leu Ile Phe Lys Asp Leu Lys Glu Leu Ser Gln
 275 280 285
 Leu Asn Leu Ser Tyr Asn Pro Ile Gln Lys Ile Gln Ala Asn Gln Phe
 290 295 300
 Asp Tyr Leu Val Lys Leu Lys Ser Leu Ser Leu Glu Gly Ile Glu Ile
 305 310 315 320
 Ser Asn Ile Gln Gln Arg Met Phe Arg Pro Leu Met Asn Leu Ser His
 325 330 335
 Ile Tyr Phe Lys Lys Phe Gln Tyr Cys Gly Tyr Ala Pro His Val Arg
 340 345 350
 Ser Cys Lys Pro Asn Thr Asp Gly Ile Ser Ser Leu Glu Asn Leu Leu
 355 360 365
 Ala Ser Ile Ile Gln Arg Val Phe Val Trp Val Val Ser Ala Val Thr
 370 375 380
 Cys Phe Gly Asn Ile Phe Val Ile Cys Met Arg Pro Tyr Ile Arg Ser
 385 390 395 400
 Glu Asn Lys Leu Tyr Ala Met Ser Ile Ile Ser Leu Cys Cys Ala Asp
 405 410 415
 Cys Leu Met Gly Ile Tyr Leu Phe Val Ile Gly Gly Phe Asp Leu Lys
 420 425 430
 Phe Arg Gly Glu Tyr Asn Lys His Ala Gln Leu Trp Met Glu Ser Thr
 435 440 445
 His Cys Gln Leu Val Gly Ser Leu Ala Ile Leu Ser Thr Glu Val Ser
 450 455 460
 Val Leu Leu Leu Thr Phe Leu Thr Leu Glu Lys Tyr Ile Cys Ile Val
 465 470 475 480
 Tyr Pro Phe Arg Cys Val Arg Pro Gly Lys Cys Arg Thr Ile Thr Val
 485 490 495
 Leu Ile Leu Ile Trp Ile Thr Gly Phe Ile Val Ala Phe Ile Pro Leu
 500 505 510
 Ser Asn Lys Glu Phe Phe Lys Asn Tyr Tyr Gly Thr Asn Gly Val Cys
 515 520 525
 Phe Pro Leu His Ser Glu Asp Thr Glu Ser Ile Gly Ala Gln Ile Tyr
 530 535 540
 Ser Val Ala Ile Phe Leu Gly Ile Asn Leu Ala Ala Phe Ile Ile Ile
 545 550 555 560
 Val Phe Ser Tyr Gly Ser Met Phe Tyr Ser Val His Gln Ser Ala Ile
 565 570 575
 Thr Ala Thr Glu Ile Arg Asn Gln Val Lys Lys Glu Met Ile Leu Ala
 580 585 590
 Lys Arg Phe Phe Phe Ile Val Phe Thr Asp Ala Leu Cys Trp Ile Pro
 595 600 605
 Ile Phe Val Val Lys Phe Leu Ser Leu Leu Gln Val Glu Ile Pro Gly
 610 615 620
 Thr Ile Thr Ser Trp Val Val Ile Phe Ile Leu Pro Ile Asn Ser Ala
 625 630 635 640
 Leu Asn Pro Ile Leu Tyr Thr Leu Thr Thr Arg Pro Phe Lys Glu Met
 645 650 655

Ile	His	Arg	Phe	Trp	Tyr	Asn	Tyr	Arg	Gln	Arg	Lys	Ser	Met	Asp	Ser
			660					665					670		
Lys	Gly	Gln	Lys	Thr	Tyr	Ala	Pro	Ser	Phe	Ile	Trp	Val	Glu	Met	Trp
		675					680				685				
Pro	Leu	Gln	Glu	Met	Pro	Pro	Glu	Leu	Met	Lys	Pro	Asp	Leu	Phe	Thr
	690					695				700					
Tyr	Pro	Cys	Glu	Met	Ser	Leu	Ile	Ser	Gln	Ser	Thr	Arg	Leu	Asn	Ser
705					710					715				720	
Tyr	Ser														